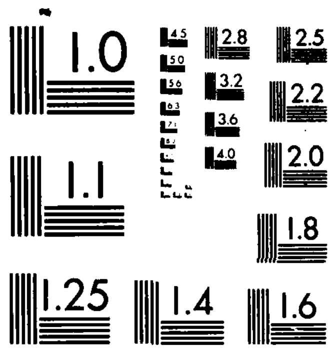
AD-A166 241 CIRCUMSCRIBING CIRCUMSCRIPTION A GUIDE TO RELEVANCE AND INCOMPLETENESS(U) MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB B C WILLIAMS OCT 85 1/1

UNCLASSIFIED AI-M-868 N00014-80-C-0505 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AIM 868	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Circumscribing Circumscription: A Guide to Relevance and Incompleteness.	5. TYPE OF REPORT & PERIOD COVERED: AI Memo
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Brian C. Williams	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0505
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Inteligence Laboratory, 545 Technology Square, Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: October 1985
	13. NUMBER OF PAGES: 46
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC
ELECTE
APR 4 1986
S D
B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

circumscription	resource limitations
commonsense reasoning	relevance
non-monotonic logic	completeness

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Intelligent agents in the physical world must work from incomplete information due to partial knowledge and limited resources. An agent copes with these limitations by applying rules of conjecture to make reasonable assumptions about what is known. Circumscription, proposed by McCarthy, is the formalization of a particularly important rule of conjecture likened to Occam's razor. That is, the set of

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(block 20 continued)

objects that is consistent with what is known.

This paper examines closely the properties and the semantics underlying circumscription, considering both its expressive power and its limitations. In addition we study circumscription's relationship to several related formalisms, such as negation by failure, the closed world assumption, default reasoning and Planner's THNOT. In the discussion a number of extensions to circumscription are proposed, allowing one to tightly focus its scope of applicability. In addition, several new rules of conjecture are proposed based on the notions of relevance and minimality. Finally a synthesis between the approaches of McCarthy and Konolige is used to extend circumscription, as well as several other rules of conjecture, to account for resource limitations.

(2)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 868 October, 1985

Circumscribing Circumscription: A Guide to Relevance and Incompleteness

Brian C. Williams

Abstract. Intelligent agents in the physical world must work from incomplete information due to partial knowledge and limited resources. An agent copes with these limitations by applying rules of conjecture to make reasonable assumptions about what is known. Circumscription, proposed by McCarthy, is the formalization of a particularly important rule of conjecture likened to Occam's razor. That is, the set of all objects satisfying a certain property is the smallest set of objects that is consistent with what is known.

This paper examines closely the properties and the semantics underlying circumscription, considering both its expressive power and limitations. In addition we study circumscription's relationship to several related formalisms, such as negation by failure, the closed world assumption, default reasoning and Planner's THNOT. In the discussion a number of extensions to circumscription are proposed, allowing one to tightly focus its scope of applicability. In addition, several new rules of conjecture are proposed based on the notions of relevance and minimality. Finally a synthesis between the approaches of McCarthy and Konolige is used to extend circumscription, as well as several other rules of conjecture, to account for resource limitations.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505.

Table of Contents

List of Figures

1. Introduction

The central concerns of logic, from the perspective of philosophy, have been the pursuit of truth and the investigation of the argumentation process. The focus of logic involves answering questions like: "What is the meaning of truth?", "How does one distinguish between valid and invalid arguments?", and "What does it mean for a statement to logically follow from another?" AI, on the other hand, is concerned with what constitutes intelligence. AI focuses on such questions as "What is the meaning of 'intelligence'?", "How does an intelligent agent perceive, reason about and interact with its environment?", "How does the agent modify its behavior based on experience?", and "How does the agent cope with its own limitations and the limitations imposed by its surroundings". Much of the formalism developed in logic to describe the meaning of truth and argumentation is applicable to AI; alone, however, this formalism is not wholly adequate. The problem is that logic in the philosophic sense is concerned with the meaning of truth independent of the limitations of any particular intelligent agent; AI, on the other hand, is very much concerned with how an intelligent agent copes with its limitations when trying to discover the truth about a matter.

One viewpoint taken in AI, and the one examined in this paper, is to view traditional logic formalisms as providing a model of an "ideal" agent who is both omnipotent and omniscient, and then to explore ways in which these logics can be modified to account for the limitations of intelligent agents in the real world. These limitations take on many forms, each resulting in an agent having incomplete or incorrect knowledge of the world. Konolige [Konolige 84], for example, discusses three forms of incompleteness resulting from reasoning based on 1) limited computational resources, 2) logically incomplete inference rules and 3) an inability to focus on relevant facts. McCarthy [McCarthy 80a], concentrates on the way in which people are able to jump to certain conclusions when faced with a situation where not all of the relevant information is available.

This paper examines the approaches taken by McCarthy and Konolige for the problem of dealing with incomplete knowledge about the physical world, focusing on circumscription and related techniques for "jumping to conclusions". The paper begins with a summary of the problems addressed and the approaches taken by McCarthy and Konolige, respectively. The remainder of the paper examines the properties of, and the relationship between, McCarthy's ideas on "simplicity" (or minimality) and Konolige's ideas on "relevance". This portion of the paper is roughly broken into two parts.

The first part begins with an informal discussion of the semantics of circumscription and its relation to techniques proposed by other researchers around the same time to cope with incomplete knowledge. This analysis leads to a generalization of circumscription that provides a means of

focusing circumscription on the relevant portion of the domain and encompasses several earlier techniques. In addition, several new techniques are proposed, providing variations on circumscription's theme of minimality. During this first part it is assumed that the reasoning agent that uses circumscription is logically complete although not omniscient and thus is not constrained by physical limitations.

The second part discusses the relationship between circumscription and Konolige's work on relevance, when applied to resource limited reasoning. This part begins with a discussion of some of the computational bottle necks involved in computing circumscription. Next, I discuss Konolige's notion of circumscriptive ignorance, a technique that allows a resource limited agent to focus on only the relevant information for a particular problem. Konolige's result is then used to extend McCarthy's circumscription, and several other rules of conjecture, to make explicit the notion of ignoring "irrelevant facts." Finally, the paper concludes with a short summary.

2. McCarthy's Circumscription

People are often faced with situations where incomplete information is specified. For example, in the missionary and cannibal problem we are presented with the task of moving two mutually antagonistic groups of people across a river, given a small boat. With only this information it is not possible to solve the problem; for example, the boat may not work correctly, due to a number of possible failures, or there may be an alternate form of transportation available, such as a helicopter, ferry, or windsurfer[1]. To qualify all the things that are not the case could require an infinite amount of information and thus is infeasible for most problems. Nevertheless people are able to solve this and similar types of problems, using only the information at hand.

To solve this problem people will jump to a number of conclusions. For example, most people assume that the boat will provide a viable form of transportation across the river unless there is any evidence to the contrary. Similarly, it is assumed that the boat is the only form of transportation since there is no evidence of any other form of transportation available. The focus of McCarthy's paper is a formalization of a particular way in which people jump to conclusions, referred to as *predicate circumscription*. According to McCarthy, predicate circumscription is a rule of conjecture that says "the objects that can be shown to have a certain property P by reasoning from certain facts A are all the objects that satisfy P". Thus by circumscribing the modes of transportation available, we believe that the only way to get across the river is by boat. In his paper, McCarthy provides a first order axiom, called the circumscription axiom, that provides a formal statement of the above intuition. In

[1] *But then again, could you picture a missionary on a windsurfer?*

addition he provides a formal semantics of predicate circumscription in terms of model theory, which he refers to as minimal entailment. Here we say that a sentence A minimally entails Q with respect to a predicate P provided Q is true in all models of A that are minimal in P. Finally, McCarthy shows how predicate circumscription subsumes an earlier form called domain circumscription which says that "the 'known' entities are all there are".

The notion of circumscription, the relationship between the circumscription axiom and minimal entailment, and the relationship between predicate circumscription and domain circumscription are all very difficult concepts to grasp. The primary objective of this paper is to explain these concepts and discuss circumscription's scope of applicability.

3. Konolige on Belief and Incompleteness

Gods are unfettered by such corporeal limitations as time and space; on the other hand, the rest of us intelligent agents are unfortunate enough to have to put up with these and other limitations. The study of logic from the viewpoint of philosophy and mathematics has been better suited for these omnipotent gods than the intelligent agents explored in AI, in that it has no way of taking into account the limitations of the physical world. Konolige identifies three types of incompleteness that result from the physical limitations of intelligent agents: 1) resource limited incompleteness, 2) fundamental logical incompleteness, and 3) relevance incompleteness. First, resource limited incompleteness occurs when ". . . an agent has the inferential capabilities to derive some consequence of his beliefs but simply does not have the computational resources to do so." Second, fundamental logical incompleteness occurs when an agent has a logically incomplete or inconsistent inference procedure.[2] Third, relevance incompleteness occurs when an agent has available all the necessary information to deduce the desired consequences, but restricts his set of knowledge in such a way that the deduction is no longer possible.

The goal of Konolige's work is to provide a formal logical framework for describing the above limitations. In accomplishing this goal Konolige's formalism differs from traditional logic formalisms in a number of ways. First, the notion of consequential closure is replaced with that of derivational closure. In addition, the rules of deduction are not required to be logically complete or sound. Thus the logic system is not guaranteed to deduce all logical consequences of what is known, but instead is only guaranteed to deduce all consequences that are *derivable* from the set of deduction rules.[3]

---

[2] It is arguable whether or not even gods are always consistent, for example, see [GODS ??].

[3] The results of derivational closure may differ from that of consequential closure when the deduction rules are logically incomplete or inconsistent.

Second, Konolige provides a framework for modeling 1) the interactions of several agents, each with possibly different limitations, and 2) the beliefs of agents about other agents (about other agents . . .). This is accomplished by allowing each agent to be modeled by a separate logical system (called a deduction structure), and then to provide a set of operators (the belief operator and the circumscription operator[4]) that allow information sharing between agents. An agent is represented by a deduction structure, consisting of 1) a set of deduction rules and 2) a set of initial beliefs in the form of sentences. The belief operator applied to an agent (A) and a sentence (P) returns true if A "believes" P. Thus if A has not deduced the truth of P or A has deduced P to be false, then the belief operator returns false. This makes it possible for agents to examine the beliefs of other agents or their own beliefs about other agents. The circumscription operator takes an agent (A), a set of sentences (L) and a sentence (P), and returns true if A can derive P from L. This makes it possible to make explicit statements about the derivation process.

Konolige's formalism provides a means of modeling several forms of incompleteness that cannot be modeled in standard first order system. This should make it possible to precisely define the semantics of several aspects of conjectural reasoning that have not yet been formalized.

The two major weaknesses of Konolige's paper are that 1) he provides little motivation for many of the components of his formalism and 2) he provides very few examples of what analytical or computational power is gained by using his formalism. It is thus very difficult to analyze Konolige's formalism on computational grounds. Instead this paper focuses on the expressive power gained by Konolige's formalism with respect to rules of conjecture similar to circumscription. See [Levesque 84] for a discussion of an alternative approach for dealing with logical incompleteness.

4. Jumping to Conclusions

Dealing with partial information is an everyday experience. People are frequently required to jump to conclusions in order to deal with a particular situation. The following are just a few typical examples.

1. The only people who said they would be going camping are John, Fred and Mary, so I'll assume the rest are not going.

2. I know my keys are here somewhere since I left them here just an hour ago.

3. There can't have been a second 'Great Depression' in 1954; otherwise, the history books would be sure to have mentioned it.

---

[4] Konolige's circumscription operator is not to be confused with circumscription in McCarthy's sense since they have little or no relation.

4. I'm sure snipes fly, they're birds aren't they?

5. He probably got all his degrees from here, after all he's the president of MIT.

6. Until recently scientists were "sure" that Saturn had exactly three rings.

We see from the above that the assumptions people make take on many forms, several of which have been investigated in AI under such names as failure by negation [Clark 78], the closed world assumption [Reiter 78], circumscription [McCarthy 80a] [McCarthy 77], default reasoning [Reiter 80] and THNOT [Sussman 70]. In the next few sections we embark upon the task of understanding the meaning behind a number of these techniques and the relationship between them. The discussion focuses primarily on the semantics of circumscription and its relationship to other techniques. The notion of circumscription is a powerful one, but one that can be very difficult to grasp. It is clear that McCarthy was only beginning to understand what circumscription was all about when he wrote about it in "Circumscription--A Form of Non-Monotonic Reasoning" [McCarthy 80a], and there was a gap of several years before other researchers understood it well enough to publish further papers on the topic. It is very difficult from McCarthy's paper to grasp the intuition behind the predicate circumscription axiom, and then to draw an exact link between this axiom and the missionary and cannibal problem discussed in the first half of his paper. Thus, before analyzing the limitations of circumscription or proposing any extensions to it, it is important to first develop an intuition behind circumscription's intended purpose. I begin by providing an informal discussion of the semantics of predicate circumscription and then show how this semantics is reflected in the circumscription axiom. A number of special cases are also considered, providing further insight into McCarthy's approach. In addition, a number of the examples listed above will be used throughout this paper as a means of comparing circumscription with other forms of conjecture.

Informally, predicate circumscription says that *the set of all objects satisfying a certain property P is the smallest set of objects that is consistent with the known facts A.*[5] For example, in the missionary and cannibal problem circumscription can be used to jump to the conclusion that the only available mode of transportation to cross the river is the boat. In this case P is the property "available modes of transportation to cross the river" and A is the fact "there is a boat available". The set {boat} is then the smallest set of "modes of transportation" that is consistent with the facts. Furthermore, by circumscribing the "ways that the boat can fail," we deduce that the boat is working correctly, since

---

[5] McCarthy describes Circumscription as "... *the objects that can be shown to have a certain property P by reasoning from certain facts A are all the objects that satisfy P.*" This statement does not convey the correct semantics of circumscription, and instead sounds almost identical to Reiter's Closed World Assumption [Reiter 78] which says roughly that the only objects that have a property P are those that logically follow from certain facts A. The difference between circumscription and the closed world assumption lies in the use of the phrase "logically consistent with", as opposed to "logically follows from". The distinction between these two techniques is discussed in detail in section 8.3.

the smallest set of ways the boat can fail that is consistent with A is the empty set.

One way of viewing predicate circumscription is as a special case of Occam's Razor. Occam's Razor says roughly that faced with several possible explanations, take the simplest one that is adequate. If we take "simplest" to mean the smallest and "adequate" to mean logically consistent with the known facts, then predicate circumscription translates to "The set of all objects satisfying a certain property P is the simplest one that is adequate." Thus the formal statement of circumscription (i.e., the circumscription axiom) provides a precise semantics for one interpretation of Occam's Razor. Having the semantics of these common sense rules of conjecture formalized is essential to further analysis and is one of circumscription's most important contributions.

5. The Semantics of Circumscription

Our informal definition of circumscription has a number of ambiguities. We say "the set of all objects satisfying a certain property P is the smallest set of objects that are consistent with the known facts A"; however, what is mean't by the words "property" and "smallest"? The objective of this section is to provide a more precise definition of each of these terms and predicate circumscription as a whole.

The use of phrases like "the set of all objects" suggests that set theory is a convenient formalism for capturing the semantics of circumscription. A "property" is taken to be a predicate on one or more individuals. A predicate P can be viewed as a set of elements, called the *extension of P*, where the elements of the extension of an n-place predicate are all n-tuples satisfying the predicate.[6] For example, if the predicate TRANSPORTATION is only true of boat, then the extension of TRANSPORTATION is the set {boat}. As a second example the natural numbers is represented by an infinite set, NATNUM $\equiv$ {0, 1, 2 . . .}.

Given a theory in terms of a set of axioms A, containing one or more instances of the predicate P, it is not always possible to determine a unique extension for P (i.e., our knowledge about P is incomplete). For example, given the axiom "TRANSPORTATION(canoe) and TRANSPORTATION(sailboat)", there is an infinite number of extensions that satisfy the predicate TRANSPORTATION, namely all sets that include both canoe and sailboat as elements. We define A_P to be the set of all extensions of P that satisfy the axioms A. Thus in the above example, taking P to be "TRANSPORTATION" and A to be "TRANSPORTATION(canoe) and TRANSPORTATION(sailboat)" we get, $A_P \equiv \{S \mid \{\text{canoe, sailboat}\} \subseteq S\}$.

---

[6] For ease of presentation we assume that all predicates are one place predicates and represent a tuple <c> simply as c; all arguments given here, however, apply to n place predicates.

Next we need to define the notion of "small". An extension is said to be *smaller* than another if it is a proper subset of the second.[7] Thus the extension {canoe, sailboat} is smaller than the extension {canoe, sailboat, windsurfer}. An extension is said to be the "smallest" if it is minimal, that is, if there is no extension that is a proper subset of it. More precisely, an extension E of a predicate P is said to be minimal with respect to a set of axioms A just in case 1) E is an extension of P (i.e., $E \in A_p$) and 2) there is no other extension of P that is smaller than E (i.e., $\neg(\exists\Phi)(\Phi \in A_p \wedge \Phi \subset E)$). Thus in the above example (where $A_p \equiv \{S \mid \{\text{canoe, sailboat}\} \subseteq S\}$), the minimal extension is {canoe, sailboat}, since none of its subsets are members of A_p. It is not necessarily the case, given a set of axioms, that a predicate has a unique minimal extension. That is, there is not necessarily a unique "smallest" set. In many cases a predicate will have several minimal extensions, and in some cases there will be none at all. Examples of both situations are provided below.

At this point we are ready to define predicate circumscription. Given a set of axioms A that only partially constrain a predicate P, predicate circumscription is a way of restricting the set of extensions of P to consist of only the minimal extensions of P in A. We call the set of extensions resulting from the circumscription of P in A *the closure of P in A* (denoted CIRC(A,P)). CIRC(A,P) is then defined to be the set of all elements S such that:

1. $S \in A_P$ and
2. $\neg(\exists\Phi.((\Phi \in A_P) \wedge (\Phi \subset S)))$

Thus CIRC(A,P), is a precise definition for one interpretation of what it means to jump to the conclusion that "the set of all objects satisfying a certain property P is the smallest set of objects which are consistent with the known facts A."

Before going on to the predicate circumscription axiom it is useful to consider a couple more examples of circumscription at the semantic level. The first example is a case where there are several minimal extensions while in the second example there are no minimal extensions at all. Consider the statement "someone left either a helicopter or a boat next to the shore", which we translate to the axiom "TRANSPORTATION(helicopter) $\vee$ TRANSPORTATION(boat)". In this case the predicate TRANSPORTATION has two minimal extensions, namely, {helicopter} and {boat}. Thus by circumscribing "modes of transportation" we deduce that "either the only mode of transportation is the helicopter or the only mode of transportation is the boat". This roughly achieves the effect of an exclusive OR, since the set {helicopter, boat} is not minimal and thus excluded as a possibility by circumscription.

---

[7] Subset is denoted $\subseteq$, proper subset is denoted $\subset$, and element of is denoted $\in$.

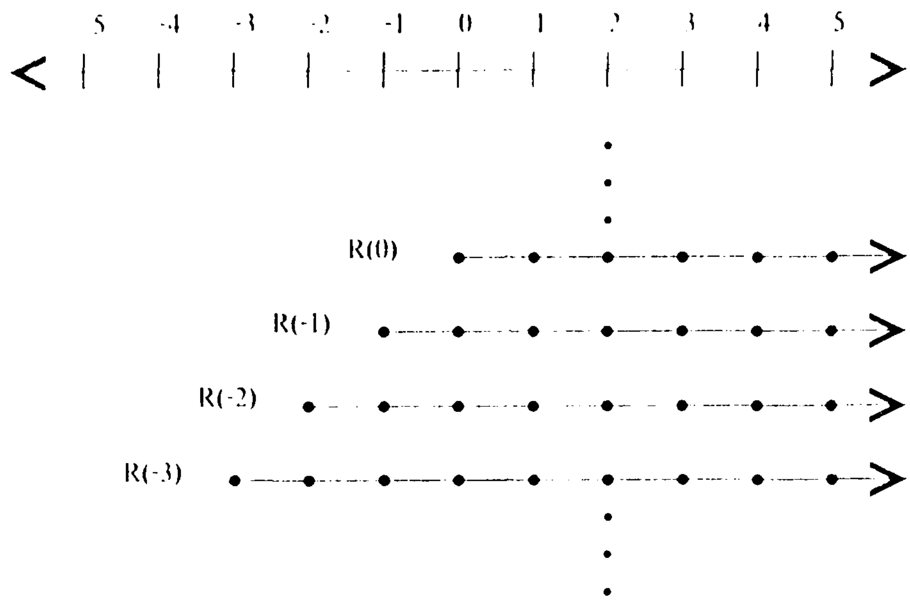


Figure 5-1: Predicate Circumscription Example: No Minimal Extension

In the above example predicate circumscription does not provide a unique minimal extension for the predicate. As a second example, consider the axiom A to be "For all integers n, P(n) implies P(successor(n))". If we define R_n to be the set of all integers greater than or equal to n, then A_P is a set containing an extension R_n for every integer n (figure 5-1). If we now arbitrarily select an extension R_n from A_P, then the extension R_{n+1} is a proper subset of R_n. Therefore, for every extension in A_P there exists another extension in A_P that is smaller, in other words there is no minimal extension of P in A. As we can see from this example, there are cases where there is no extension of P that satisfies circumscription; thus it is sometimes the case that circmscribing a theory will make the theory *inconsistent*.

6. The Predicate Circumscription Axiom

In the previous section I defined the result of circumscribing a predicate P with respect to a theory A as the predicate CIRC(A,P), described extensionally as:

$$\text{CIRC}(A,P) \equiv \{S \mid S \in A_P \wedge \neg(\exists \Phi.((\Phi \in A_P) \wedge (\Phi \subset S)))\}$$

Where A_P is the set of all extensions of P that satisfy the axioms A.

To achieve the results of the CIRC operator in the proof theory of first order logic, for example, it is necessary to augment the set of axioms of a theory A in such a way that all the extensions of P in A are minimal. The predicate circumscription axiom is a means of achieving this effect. The goal of this

section is to show how the circumscription axiom together with the set of axioms A implies that P is equivalent to CIRC(A,P).

For the purpose of this discussion we take the circumscription axiom to be the following second order sentence schema quantifying over all predicates Φ.

Definition *The circumscription of P in A is the sentence schema*[8]

$$\forall\Phi.(((\Phi \in A_p) \land \forall x.(\Phi(x) \supset P(x))) \supset \forall x.(P(x) \supset \Phi(x))). \tag{1}$$

In the above definition the first conjunct in the antecedent, $\Phi \in A_p$, is a predicate on predicates which is true exactly when every extension of Φ in A is also an extension of P in A.[9] Thus the first conjunct says that for each model of the theory the extension of P is an element of A_p.

The second conjunct of the definition, $\forall x.(\Phi(x) \supset P(x))$, says that P is true of an individual whenever Φ is. Thus, in a particular model of the theory, every element of the extension of Φ is also a member of the extension of P, or equivalently Φ is a subset of P ($\Phi \subseteq P$). Similarly the antecedent of the definition, $\forall x.(P(x) \supset \Phi(x)))$, is taken to mean P is more specific ($P \subseteq \Phi$). Making these substitutions the circumscription axiom becomes:

$$\forall\Phi.(((\Phi \in A_p) \land (\Phi \subseteq P)) \supset (P \subseteq \Phi))$$

Another way of viewing this statement is that, for each model of the theory, *any predicate Φ that is both a member of A_p and a subset of P must be equivalent to P.*

$$\forall\Phi.(((\Phi \in A_p) \land(\Phi \subseteq P)) \supset(P \equiv \Phi))$$

Or equivalently *there exists no predicate Φ which is an element of A_p and which is a strict subset of P.*

$$\neg(\exists\Phi.((\Phi \in A_p) \land(\Phi \subset P))) \tag{2}$$

This is exactly the second part of the definition of CIRC, (where P is taken to be S in the definition).

Next, given the set of axioms A, the first half of the definition trivially follows since the extensions of P are necessarily a subset of the extensions of P. (Syntactically this is equivalent to replacing all instances of P in A with itself, producing just A.) Thus for each model of the theory:

$$P \in A_p \tag{3}$$

Finally, from equations (2) and (3), and the definition of CIRC, we deduce that the set of extensions of

---

[8] Again, the circumscription axiom schema is easily generalized to handle arbitrary n place predicates by replacing x with an n-tuple.

[9] The term $\Phi \in A_p$ is constructed in a first order axiom by substituting all instances of P in A by Φ (i.e. $\Phi \in A_p \equiv \lambda\Phi.P\,S^P_\Phi A$).

P, resulting from the set of axioms A *plus* the circumscription axiom, is the set of all minimal extensions of P in A. Thus the circumscription axiom captures the semantics described in the previous section.

7. Properties of the Circumscription Axiom

Given an axiom that captures the meaning of circumscription, what do we do with it? The objective of this section is to answer this and similar questions. This section begins with a discussion of the general properties of circumscription and a number of ways it can be used to make deductions. The discussion then turns to several ways that circumscription is used in the common sense world. Finally, circumscription's limitations are discussed, providing motivation for later sections.

One of the most important contributions of circumscription is that it provides a precise statement of the semantics of an important form of conjectural reasoning. This statement of circumscription is useful in analyzing both its computational properties and its expressive power. In this section we are primarily interested in circumscription's expressiveness; its computational properties are examined in the second half of this paper. Two ways of using circumscription are examined below -- the first consists of determining the set of all individuals that satisfy a circumscribed predicate, while the second is a way of determining whether or not a circumscribed predicate is true of a particular individual.

One of the appealing properties of circumscription is that it works within the framework of first order logic, whose properties are fairly well understood, rather than creating a new logic that hasn't yet been characterized. To incorporate circumscription into a first order system, the circumscription axiom is converted from a second order statement to a first order axiom schema, by removing the quantification over Φ, and instead viewing Φ as a predicate parameter for which an arbitrary expression can be substituted.

$$((\Phi \in A_p) \wedge \forall x.(\Phi(x) \supset P(x))) \supset \forall x.(P(x) \equiv \Phi(x)). \quad (4)$$

7.1. Determining the Individuals That Satisfy a Predicate

The need for circumscription arises from the need to reason based on incomplete information. Basically, given an incomplete description of the set of individuals satisfying a particular property, circumscription is an intuitively satisfying assumption about how to complete this set. What the circumscription axiom provides us with is a precise way of stating this assumption. Given a predicate P and a set of axioms that we want to circumscribe over, one way the circumscription axiom is typically used is to determine the set of all individuals that satisfy the predicate. To accomplish this, one first constructs a predicate Φ describing a set of individuals and then uses the circumscription

axiom to show that Φ is equivalent to P. (where P is the circumscribed predicate we're interested in). This, in turn, is accomplished by instantiating the circumscription axiom schema (equation 4) with a specific A, P and Φ, and then showing that the antecedent of the axiom follows from what is known. Thus the major steps in using circumscription are 1) selecting a predicate P to be circumscribed, 2) selecting a set of axioms, A, to circumscribe over, 3) generating Φ, and 4) showing that the antecedent of the instantiated axiom follows from what we know. For example, consider the circumscription of P in A, where P is the predicate "TRANSPORTATION" and A is the sentence "TRANSPORTATION(canoe) $\wedge$ TRANSPORTATION(sailboat)". We then guess canoe and sailboat to be the only modes of transportation and instantiate the circumscription axiom schema with $\Phi(x)$ being the expression "(x = canoe) $\vee$ (x = sailboat)":

$$\begin{aligned}&((((\text{canoe} = \text{canoe}) \vee (\text{canoe} = \text{sailboat})) \wedge ((\text{sailboat} = \text{canoe}) \vee (\text{sailboat} = \text{sailboat}))) \\ &\wedge \forall x.(((x = \text{canoe}) \vee (x = \text{sailboat})) \supset \text{TRANSPORTATION}(x))) \\ &\supset \forall x.(\text{TRANSPORTATION}(x) \equiv ((x = \text{canoe}) \vee (x = \text{sailboat}))).\end{aligned} \tag{5}$$

The first part of the antecedent in (5) is tautologically true while the second follows from A; thus, from the consequent of (5), the minimal set of transportation modes is {canoe, sailboat}. In this example, the circumscribed predicate is described by a unique minimal extension and thus is completely determined.

If a predicate P has several minimal extensions, then the result of circumscribing P may still provide useful information. For example, suppose that TRANSPORTATION is being circumscribed in the sentence "(TRANSPORTATION(canoe) $\wedge$ TRANSPORTATION(sailboat)) $\vee$ (TRANSPORTATION(canoe) $\wedge$ TRANSPORTATION(kayak))" using the above technique we can show that:

$$\begin{aligned}&\forall x.(\text{TRANSPORTATION}(x) \equiv ((x = \text{canoe}) \vee (x = \text{sailboat}))) \\ &\vee \forall x.(\text{TRANSPORTATION}(x) \equiv ((x = \text{canoe}) \vee (x = \text{kayak})))\end{aligned} \tag{6}$$

Thus TRANSPORTATION has two minimal extensions, {canoe, sailboat} and {canoe, kayak}. From this we can deduce several things; for example, in either case there are only two modes of transportation available, one being a canoe and both being water vehicles. Using McCarthy's terminology, we say that equation (6) minimally entails each of these facts with respect to TRANSPORTATION, since each fact holds in all minimal extensions of TRANSPORTATION. This differs from regular entailment, since a fact may be true in all models A where P is minimal, and yet not be true in all models of A. For example, there are exactly two modes of transportation in all minimal extensions of TRANSPORTATION in (6); however, there are extensions of TRANSPORTATION in (6) where there are more than two modes of transportation. The topic of minimal entailment arises again later in this paper during the discussion of non-monotonic reasoning (section 11).

Also consider the case where P has no minimal extensions in A (an example of this was provided at

the end of section 4). If we use the second order statement of the circumscription axiom (equation (1)) then there is no extension of P (and therefore no model) such that A and equation (1) both hold. Thus A taken together with the circumscription axiom applied to P is *inconsistent*! Fortunately the effect this has on the first order circumscription axiom schema is that there will be no instantiation of Φ with P such that the consequent of the circumscription axiom can be deduced. Thus at worst the circumscription axiom will provide no useful information.

7.2. Determining the Truth of a Predicate for a Single Individual

The above discussion described how the circumscription axiom is used to determine all the minimal extensions of a predicate P. This is similar to the way circumscription is used in McCarthy's paper. As is discussed in later sections, finding an instantiation for Φ that enumerates all minimal extension of P can be computationally quite expensive. Searching through all possible instantiations of the predicate Φ is equivalent to searching through the space of all possible formulas. This section proposes a second technique, not discussed in the literature, that uses the circumscription axiom to determine if P is true or false of a particular individual, a, in all minimal extensions of P, while avoiding the computational cost of the technique described above.

An obvious way of accomplishing the above task is to find all minimal extensions of P and then test the truth of P(a) in each extension. This approach is undesirable since it essentially involves determining the truth of P for every individual in the domain. A much more desirable way of determining the truth of P(a) is to find an instantiation of Φ that only constrains the truth of Φ applied to a, while leaving the truth of Φ applied to all other individuals to be the same as that of P. This avoids the work of unnecessarily determining the truth of P for all other individuals in the domain. For example, P(a) can be shown to be false in all minimal extensions of P in A by instantiating Φ(x) as "(x ≠ a) ∧ P(x)" in the circumscription axiom (4), and then showing the antecedent of the axiom to be true. Performing this substitution, the circumscription axiom schema simplifies greatly, becoming:

$$(((x \neq a) \wedge P(x)) \in A_p) \supset \neg P(a). \tag{7}$$

Consider the previous example where P is TRANSPORTATION and A is "(TRANSPORTATION(canoe) ∧ TRANSPORTATION(sailboat)) ∨ (TRANSPORTATION(canoe) ∧ TRANSPORTATION(kayak))". Circumscribing P in A we might want to show that there are no helicopters available (i.e., ¬TRANSPORTATION(helicopter)). Using (7), and instantiating the individual, a, as helicopter:

$$\begin{aligned}
&(((\text{canoe} \neq \text{helicopter}) \wedge \text{TRANSPORTATION}(\text{canoe}) \\
&\quad \wedge (\text{sailboat} \neq \text{helicopter}) \wedge \text{TRANSPORTATION}(\text{sailboat})) \\
&\;\vee ((\text{canoe} \neq \text{helicopter}) \wedge \text{TRANSPORTATION}(\text{canoe}) \\
&\quad \wedge (\text{kayak} \neq \text{helicopter}) \wedge \text{TRANSPORTATION}(\text{sailboat}))) \\
&\supset \neg\text{TRANSPORTATION}(\text{helicopter}).
\end{aligned}$$

In the above equation, all of the inequalities are true and the antecedent simplifies to A. therefore, ¬TRANSPORTATION(helicopter) follows. We have thus determined the truth of TRANSPORTATION(helicopter), while avoiding the expense of finding a "full instantiation" of Φ as was performed in the first technique.

The above example demonstrates a way of showing that P(a) is false in all minimal extensions of P in A. The second case involves showing that P(a) is true in the same situation. To deal with this case, note that if a is a member of a minimal set describing P then it will also be a member of all supersets of P. Furthermore, each extension of P is a superset of some minimal extension of P. Thus if P(a) is true in all minimal extensions of P then it is also true in all extensions of P. Therefore, A minimally entails P(a) with respect to P if A entails P(a). In other words, nothing is gained from circumscription in trying to prove the truth of P(a)), and the circumscription axiom reduces to $P(a) \supset P(a)$.

7.3. Applying Circumscription and Other Forms of Conjecture

When using circumscription, one must select both a predicate and a set of axioms to be circumscribed over. The set of axioms selected depends on properties of both the domain and the intelligent agents being modeled. If the intelligent agent is omnipotent, then it will use all the available information, and thus circumscribe over everything that is known.[10] On the other hand, if the agent is resource limited, then a subset of the known facts are selected that are assumed to be relevant to the problem.

The problem of selecting a predicate, however, plagues both resource limited and omnipotent agents alike. In the examples that have been considered thus far, the choice of a predicate to be circumscribed has been obvious. There are, however, many cases in which the choice is not so obvious. Consider the following blocks world example, where the domain consists of the 6 blocks a through f, the predicates BLACK and WHITE, and the knowledge that every block is BLACK or WHITE (i.e., $\forall x.(\mathrm{BLACK}(x) \equiv \neg\mathrm{WHITE}(x))$). In addition it is given that blocks a and b are BLACK, while c and d are WHITE. If we wanted to determine what blocks were BLACK we could use circumscription to deduce that the set of all BLACK blocks is {a,b}. It then follows that the remaining blocks {c,d,e,f} are WHITE. On the other hand, by circumscribing the predicate WHITE it follows that {c,d} are WHITE, while {a,b,e,f} are BLACK.

We see from this example that circumscription can inadvertently produce a number of undesirable side effects. These side effects can become quite subtle in complex systems. One ramification of this is that one must be careful in determining whether to circumscribe a predicate or its negation. It is

---

[10] This appears to be an implicit assumption in McCarthy's paper, based on the examples he provides.

important that one makes a conscious decision about which of the two is being circumscribed. A second ramification is that the result of multiple circumscriptions is order sensitive. This is a serious logical problem; the ordering of predicates to be circumscribed is a result of the linearity of syntax and has little semantic significance. The fact that successive circumscriptions are not "associative" means that one must make arbitrary decisions about the order of the circumscribed predicates.

To deal with these issues one must consider the properties of the problem domain to which circumscription is being applied. We now consider a number of ways that circumscription and similar forms of conjectural reasoning have been used, including in database systems, common sense reasoning and computer-aided instruction. The relationship between circumscription and several other rules of conjecture discussed below is the topic of later sections. In addition, a solution to the problem of circumscription's ordering sensitivity is discussed in section 8.2.

A number of rules of conjecture, such as *failure by negation* [Clark 78] and *the closed world assumption* [Reiter 78], arose out of work on database systems. One of the major issues in dealing with large databases is the *negative information problem.* In general the amount of negative information is often much too large to represent explicitly in the database. [Naqvi 85] For example, in an airline flight reservation system it would be much too unwieldy to represent all the places that an airline cannot fly to, as well as the places it can. Such a data base, which represents both positive and negative information explicitly, is referred to as an *open database*. On the other hand, in a *closed database* only positive information is represented explicitly, while the truth of negative facts is deduced by default. More specifically, any fact that doesn't logically follow from the set of facts in the database is false. Thus, in the above example, there is no flight route between two cities unless one can be deduced. Negation by failure and the closed world assumption are both techniques for making this type of deduction. A closed database is a convention of communication that makes it possible to represent information concisely. This convention appears in many areas, such as the representation of circuit connections with a schematic or routes with a road map.[11]

As discussed above, circumscription plays a very similar role in the area of common sense reasoning. For example, if I'm asked who is going camping, and I respond that John, Fred and Mary are, then it is assumed by convention that they are the only people that are going. If this isn't the case then I would qualify my answer with something like "to the best of my knowledge" or "the people I know".

---

[11] One thing that is lost with such a convention is that it is no longer possible to represent a piece of knowledge being unknown since anything that cannot be shown to be true is assumed to be false. Thus the implicit representation of negative facts presumes total knowledge about the domain being represented.

An important difference between the use of these techniques in formal database systems versus common sense domains is the certainty with which a negative fact is held. In a formal database application, closure is an explicitly stated convention that is held with great certainty. Thus in these applications it is very important that the integrity of the database is assured. On the other hand, in a common sense domain the negative information is taken only as the most likely answer given what is known thus far. For example, it was assumed until recently that Saturn had only three rings. Furthermore, this fact often appeared in school texts without any qualification. Thus the discoveries of the Voyager missions required a major revision of the public's beliefs. This type of conjectural reasoning occurs over and over again in science. Someone comes up with a hypothesis that is either refuted or becomes stronger and stronger as the evidence collects. In computer science, the class of NP-complete problems is an explicit attempt to accumulate evidence to support a belief that "P does not equal NP".

The problem of measuring the certainty to which we hold a belief, is an interesting and difficult problem. In problems of scientific investigation this certainty might be derived empirically. In other more common sense domains our certainty might be based on other factors, such as our model of the person we are communicating with. One approach, investigated by Collins and his colleagues [Collins 75] is based on the importance of a particular fact. For example, given a question like "were any U.S. Presidents women?", their system would reason that 1) it knows of no women presidents, and 2) the fact is sufficiently important that the system would have heard of it if it was true, thus the answer must be false. On the other hand, given a question like "was it a good year for raspberry picking in Oakland County, Michigan?", the system would probably fail to answer on the basis of lack of information (since the fact isn't important enough that the system would have any reason for knowing it). Of course this research opens as many questions as it answers. For example, how does one determine whether or not a fact is important? Do people place different levels of importance on facts? What is our confidence in a person telling us the relevant information for a problem? Many of these questions are intimately wrapped up in our model of belief. Several attempts have been made to formalize the notion of "belief" [Doyle 80], [Weyhrauch 80], including Konolige's article "Belief and Incompleteness" [Konolige 84] discussed later in this paper.

Above I have discussed several uses of both circumscription and conjectural reasoning in general. During this discussion a number of issues were raised that are as yet unsolved in the current research. In addition, many of the above examples involving other forms of conjectural reasoning cannot be performed using circumscription as it stands. Several of these limitations are discussed in the next three sections, along with proposals for extending circumscription. The first section discusses adding to circumscription the ability to focus in on particular portions of the domain to be

circumscribed. The second talks about how to incorporate defaults into circumscription, and the third discusses circumscribing over things other than predicates.

8. Relevance: Tuning the Scope of Predicate Circumscription

Roughly speaking, the predicate circumscription axiom provides a means of stating in first order logic that the set of individuals in the domain that have a particular property is the smallest set that is consistent with what is known. In some cases we would like to jump to certain conclusions about a property without having it apply to all individuals in the domain. For example, we might want to say that the only birds that don't fly are those that are known, without making any commitment about flying mammals or fish. On the other hand, often we would like to expand the scope of circumscription so that it applies to several properties at once. Thus we need a way of specifying the set of relevant individuals and properties that circumscription is being applied to. This is the topic of the first two parts of this section. If we expand the scope of circumscription to encompass all predicates and individuals in the theory we have something similar (but not equivalent) to the closed world assumption. A comparison between circumscription and the closed world assumption is the topic of the third part of this section.

8.1. Moving In - Focusing on the Relevant Individuals

In the above example we want to circumscribe the predicate FLIGHTLESS over the set of individuals that are birds, without making commitments about any other type of individual. A straight forward application of the circumscription axiom, however, results in circumscribing over *all* individuals in the domain, clearly not what we desire. Instead we would like to specify a subset of the domain and circumscribe over it, while leaving the rest of the domain untouched. The subset of the domain we're interested in can be represented as a characteristic function C(x), which is simply a predicate on the individuals of interest. Thus C(x) can be viewed as a subset of the individuals in the domain, just like any other predicate.

To circumscribe a predicate P in A over only the individuals in C, we need to remove the individuals in P we are not interested in and then minimize over the resulting set. To accomplish this we intersect both P and Φ with C, to remove the undesirables, and then use the definition of minimal, giving us:

$$\neg(\exists\Phi.((\Phi \in A_p) \wedge ((\Phi \cap C) \subset (P \cap C)))) \qquad (8)$$

The corresponding second order axiom is then:

$$\forall\Phi.(((\Phi \in A_p) \wedge \forall x.((\Phi(x) \wedge C(x)) \supset (P(x) \wedge C(x)))) \\ \supset \forall x.((P(x) \wedge C(x)) \equiv (\Phi(x) \wedge C(x)))).$$

which simplifies to:

$$\forall\Phi.(((\Phi \subset A_p) \wedge \forall x.((\Phi(x) \wedge C(x)) \supset P(x))) \supset \forall x.((P(x) \wedge C(x)) \supset \Phi(x))). \quad (9)$$

Thus in the bird example, we can say "the only birds that are flightless are the ones we know about" by letting P be FLIGHTLESS and C be BIRD. If, for example, we knew only that "a penguin is a flightless bird," ($A \equiv$ (BIRD(penguin) $\wedge$ FLIGHTLESS(penguin))) then by instantiating Φ in axiom (9) as:

$$\Phi(x) \equiv (\text{BIRD}(x) \wedge (x = \text{penguin})) \vee (\neg\text{BIRD}(x) \wedge \text{FLIGHTLESS}(x))$$

$\Phi \subset A_p$ is a tautology, and thus it follows that:

$$\forall x.((\text{BIRD}(x) \wedge \text{FLIGHTLESS}(x)) \supset (x = \text{penguin}))$$

In addition we can use axiom (9) to test whether or not a particular bird is flightless using the technique developed earlier for predicate circumscription.

Next consider the limiting cases for the characteristic function C(x). If we take C to be the complete domain of individuals (i.e., $\forall x.C(x)$), then axiom (9) becomes logically equivalent to the predicate circumscription axiom, just as we would expect. At the other extreme if we take C to be a single individual, a, (i.e., $C(x) \equiv (x = a)$) then axiom (9) becomes:

$$\forall\Phi.(((\Phi \in A_p) \wedge (\Phi(a) \supset P(a))) \supset (P(a) \supset \Phi(a)))$$

which says "assume that P(a) is false as long as it is consistent with A". This is equivalent to the negation by failure inference rule, which states that " . . . $\neg$ P can be inferred if every possible proof of P fails." [Clark 78] Thus far we have described how to circumscribe over any subset of the individuals in the domain. The next section describes how to circumscribe over any subset of the *predicates* in the domain.

8.2. Moving Out - Focusing on the Relevant Properties

An obvious way of applying circumscription to multiple predicates is to instantiate the circumscription axiom sequentially on each predicate. For example, to circumscribe predicates P and Q over a set of axioms A, one would first instantiate equation (4) with P and A, and then circumscribe Q over the resulting set of axioms. However, recall from the blocks world example above that circumscription is order sensitive; circumscribing WHITE and then BLACK produces a different result than circumscribing the predicates in the opposite order. In practice, when a person decides to circumscribe their knowledge of the world for a particular problem, the circumscription is applied over a set of properties. This is an important point: *it is a set, not a sequence that we are circumscribing over.* Thus the ordering of these properties has no semantic import and should be irrelevant to the circumscription. Instead the circumscription axiom must be augmented so that several circumscriptions occur simultaneously, avoiding the problem of ordering sensitivity. McCarthy suggests a generalization of the predicate circumscription axiom that does this. For the

case where two predicates, P and Q, are being jointly circumscribed, the axiom becomes:

$$((\langle\Phi,\Psi\rangle \in A_{P,Q}) \wedge \forall x\,(\Phi(x) \supset P(x)) \wedge \forall x.(\Psi(x) \supset Q(x))) \quad (10)$$
$$\supset (\forall x.(P(x) \equiv \Phi(x)) \wedge \forall x.(Q(x) \equiv \Psi(x))).$$

Viewing $A_{P,Q}$ as a set of pairs of extensions for P and Q respectively, then:

$$\neg(\exists\langle\Phi,\Psi\rangle.((\langle\Phi,\Psi\rangle \in A_{P,Q}) \wedge (\langle\Phi,\Psi\rangle \neq \langle P,Q\rangle) \wedge (\Phi \subseteq P) \wedge (\Psi \subseteq Q))) \quad (11)$$

Thus equation (10) naturally extends the notion of minimality to encompass two predicates.[12] Generalizing to any number of predicates, the axiom states that, for each model of the theory, *there exists no tuple $\langle\Phi,\Psi, \ldots\rangle$ in $A_{P,Q,\ldots}$ which is distinct from $\langle P,Q, \ldots\rangle$, and whose elements are subsets of the respective elements in $\langle P,Q \ldots\rangle$.*[13]

Returning to the circumscription of the two predicates P and Q, if P and Q are independent in A,[14] then the result of circumscribing P and Q simultaneously is equivalent to that of circumscribing P and Q separately (in either order). In the blocks world example of the previous section, if we remove the constraint that "every block is BLACK or WHITE," it then follows from circumscribing BLACK and WHITE together that {a,b} are BLACK, {c,d} are WHITE, and {e,f} are neither. This is equivalent to circumscribing WHITE alone, followed by BLACK (and vice-versa). On the other hand, including the constraint between BLACK and WHITE, the result of circumscribing the predicates together is:

$$\langle \text{BLACK}, \text{WHITE}\rangle \equiv \langle\{a,b\},\{c,d,e,f\}\rangle \vee \langle\{a,b,e\},\{c,d,f\}\rangle$$
$$\vee \langle\{a,b,e,f\},\{c,d\}\rangle \vee \langle\{a,b,f\},\{c,d,e\}\rangle$$

This result includes both the results of circumscribing BLACK followed by WHITE and WHITE followed by BLACK as subsets as well as results not found in either.

The technique described in this section allows us to expand the scope of circumscription to include several predicates in the domain at once. Similarly, the technique of the previous section provides a means of focusing on any subset of the individuals in the domain. Thus taken together, these two techniques provide a powerful tool for selecting the relevant slice of the domain to be circumscribed. These are only two of the ways that the predicate circumscription axiom can be constrained. Other ways of constraining predicate circumscription are explored in later sections. In the previous section we considered the limiting case where a predicate is circumscribed over a single individual. The next section considers the other limiting case where the whole domain is circumscribed.

---

[12] Recall that the minimality condition for a single predicate P is equivalent to:

$$\neg(\exists\Phi\,((\Phi \in A_P) \wedge (\Phi \subseteq P) \wedge (\Phi \neq P)))$$

[13] A second way of viewing the circumscription of P and Q in A is that *the set of terms involving P or Q that are true in A, is minimal*.

[14] We say that P and Q are independent in A if $A_{P,Q} = A_P \times A_Q \equiv \{\langle p,q\rangle \mid (p \in A_P) \wedge (q \in A_Q)\}$

8.3. The Limiting Case: Predicate Circumscription and the Closed World Assumption

By extending circumscription to close over all predicates in A, we capture the intuition of the conjecture that "the set of things we believe to be true is the minimal set that is consistent with what we know" (where, in this case a "thing" is taken to be a term). This is very similar to Reiter's closed world assumption (CWA) [Reiter 78], which roughly states that "we only believe those things to be true that follow from what we know". Although these statements might appear the same at first glance, they have a subtle, but important, difference. This difference arises from the use of the phrases "consistent with" in circumscription, versus "follows from" in the CWA, and is best seen through an example. Consider once again the missionary and cannibals problem where we are told that either a boat or a helicopter is available. In this example the result of circumscribing TRANSPORTATION was two minimal extensions, {helicopter} and {boat}. Next consider the application of the CWA to the same example. It doesn't logically follow from the statement "TRANSPORTATION(helicopter) $\vee$ TRANSPORTATION(boat)" that helicopter is an available mode of transportation; neither does this follow for a boat. Thus, the result of applying the CWA is that there are no available modes of transportation to cross the river, but this is clearly inconsistent with the original statement that either a helicopter or a boat is available!

Let's examine this inconsistency more carefully for a moment. Another way of stating the CWA is that a literal[15] is assumed to be false as long as the assumption is consistent (i.e., it doesn't follow that the literal is true). Thus the CWA checks to make sure that each assumption taken separately is consistent with the original data base. It does not, however, check to make sure that the assumptions are *mutually* consistent, thus inconsistencies are allowed to slip by. Circumscription differs in that it makes precisely this check by assuring that the predicates extensions be both minimal and consistent.

Reiter recognizes this inconsistency, and avoids it by restricting the applicability of the CWA to horn databases. A horn database consists of a set of clauses, each of which has at most one positive literal. Each clause can be viewed as an implication where the consequent consists of the positive literal and the antecedent consists of all the negative literals. For example, the following set of clauses are horn:

$$(\neg \mathrm{SNIPE}(x) \vee \mathrm{BIRD}(x))$$
$$\wedge(\neg \mathrm{PENGUIN}(x) \vee \mathrm{BIRD}(x))$$

and are equivalent to:

$$\forall x.((\mathrm{SNIPE}(x) \vee \mathrm{PENGUIN}(x)) \supset \mathrm{BIRD}(x))$$

---

[15] A positive literal is a term (a predicate applied to a tuple of individuals), while a negative literal is a negated term.

A way of applying the CWA, or "closing off the database" for a literal $P(x)$ is suggested by Clark [Clark 78], which consists of 1) collecting all the clauses whose consequents contain $P(x)$, 2) converting the clause to the form: $\forall x.(A(x) \supset P(x))$, and then 3) inverting the implication and combining with the original clause to get $\forall x.(A(x) \equiv P(x))$. This is equivalent to saying that the sufficient conditions for the positive literal to be true are also the necessary ones.

Given the above restrictions, Reiter claims in [Reiter 82] that Clark's completion axiom, just described, is implied by predicate circumscription. In his paper, Reiter provides an example where this is the case, and states this implication as a theorem; however, he neglects to provide any proof of the theorem. Furthermore, this proof does not seem to appear in any of his other published works. Reiter then implies at the end of the paper that, as a result of this theorem, predicate completion can be used to generate an instantiation of $\Phi(x)$ that is equivalent to CIRC(A,P). If true, this would be an important result since it would remove the guess work involved in finding the minimal predicate. However, this interpretation appears to false. Consider the following example, consisting of the domain of integers and the single axiom:[16]

$$\forall n.((n = 0 \lor P(n - 1)) \supset P(n)) \qquad (12)$$

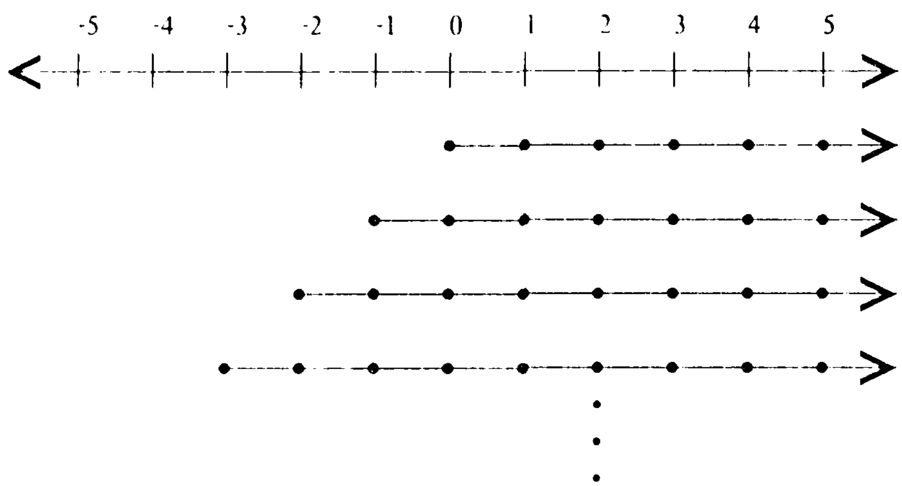


Figure 8-1: Upwardly Closed Rays That Include Zero

Semantically this axiom describes the set of all P's, each of which is upwardly closed and includes zero (figure 8-1). This axiom is in the desired form, $\forall x.A(x) \supset P(x)$, prescribed by Clark for predicate completion, thus we infer that:

$$\forall n.(P(n) \supset (n = 0 \lor P(n - 1))) \qquad (13)$$

---

[16] This axiom is equivalent to the conjunction of the following two horn clauses:

$\forall n.((n = 0) \supset P(n))$

$\forall n.(P(n - 1) \supset P(n))$

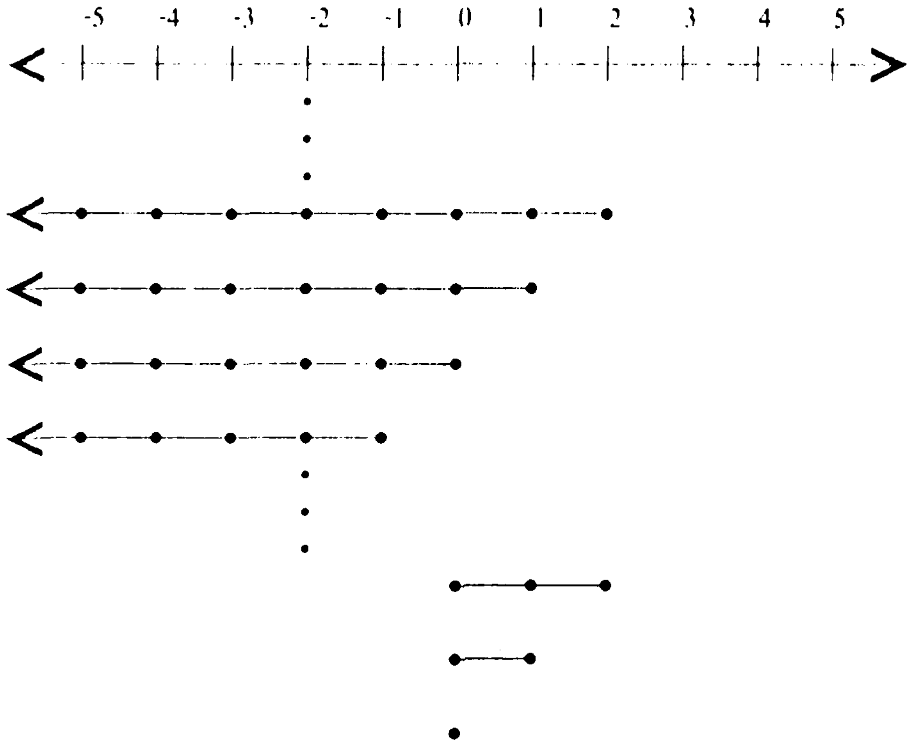


Figure 8-2: Rays That are Downwardly Closed Except at Zero

This is equivalent to a set of P's, each of which is downwardly closed except at $n = 0$ (figure 8-2). As a result of predicate completion the conditions of (13) are combined with (12), providing two extensions for P, the first containing exactly the natural numbers and the second containing the integers (figure 8-3). The first case is a minimal extension of P, and is the same as the result of applying circumscription. The second case, however, is clearly not minimal, since the integers is a strict superset of the natural numbers. Thus using predicate completion to perform the closed world assumption does not necessarily result in constraining P's extensions to be minimal. In other words, the results of predicate circumscription and the closed world assumption are not equivalent.

This example provides a counter example to Reiter's interpretation of the theorem, that is, that predicate completion can be used to find an instantiation of the predicate that is minimal for horn databases. What Reiter's theorem is saying is something much weaker. That is, for horn databases, the predicate resulting from predicate completion is true in all cases that the circumscribed predicate is true. Thus predicate completion is not over restrictive with respect to circumscription. This is a useful result since it tells us that, like circumscription, the CWA will not result in an inconsistency when applied to horn data bases (earlier we discussed how in the more general case the CWA can produce inconsistencies). However, the theorem says nothing about predicate completion being

under restrictive, that is, how close the predicate resulting from predicate completion is to being minimal. In fact, what the above example shows is that in some cases the results of predicate completion can be far from minimal! Thus it is not clear what benefit we have gained in using the CWA over leaving the predicate completely alone. Certainly the uncircumscribed (or uncompleted) predicate would also be guaranteed not to be over restrictive. Intuitively (and in practice) the CWA appears to be quite useful as a circumscriptive device. However, it remains to be shown exactly how restrictive the CWA really is.

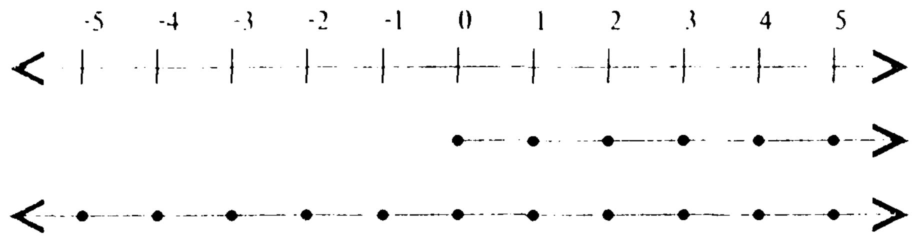


Figure 8-3: Combination of the Upwardly and Downwardly Closed Rays

Another problem with the CWA is that, like circumscription, it is ambiguous whether or not a predicate or its negation should be assumed. Recall that the CWA says that a positive literal is assumed to be false unless it logically follows that it is true. On the other hand, the fact that we need to make an assumption implies that it does not logically follow that the literal is false. Therefore, it is logically consistent to assume either that the literal is false or true. The CWA always assumes that the truth of a positive literal is false unless it follows that it is true. The default assumption for the truth of a literal, thus, *is built implicitly into how the literal is expressed*. For example, when talking about birds and flight we might select the predicate FLIES and list birds that can fly or select the predicate FLIGHTLESS and list those birds that can't. The selection of the particular predicate can have a dramatic affect. For example, by selecting FLIES it is assumed that the only birds that can fly are those that are mentioned, while selecting FLIGHTLESS assumes that the ones mentioned are the only ones that can't fly. These types of decision are particularly important during conjectural reasoning, since statements like "penguin's can't fly" can suddenly become the norm, rather than the exception. One of the motivations for default reasoning is to make these assumptions explicit in the axioms. The next section discusses a way of incorporating defaults into circumscription.

9. Circumscription and Default Reasoning

The types of conjectural reasoning examined thus far say something roughly like "the objects that can be shown to have a certain property P by reasoning from certain facts A are all the objects that satisfy P". On the other hand, the notion of a default, P, is something like "the objects that can be

shown not to have a certain property P by reasoning from certain facts A are the only exceptions". Thus a default is a property that is true of a class of individuals. Default reasoning is prevalent in many domains, such as knowledge representation, temporal reasoning, troubleshooting, qualitative reasoning and so forth. In knowledge representation we use defaults to say something like "birds usually can fly" and "elephants usually are gray, have four legs and a trunk." In temporal reasoning defaults are embodied in the persistence assumption, "things don't change without a cause". To perform default reasoning, one must provide a way of 1) stating a default property for a class of individuals and 2) determine whether or not a particular individual satisfies a default. This section builds on the results of previous sections to incorporate default reasoning into circumscription.

The notion of a default is analogous to circumscription. The difference is that default reasoning tries to *maximize* the number of objects having a particular property, while circumscription tries to *minimize* it. Thus default reasoning can be restated as *the set of all objects satisfying a certain property P is the largest set of objects that are consistent with the known facts A.* That is, default reasoning is a way of restricting the set of extensions of P to consist of only the *maximal* extensions of P in A. An axiom for default reasoning, analogous to the predicate circumscription axiom is:

$$\forall\Phi.(((\Phi \in A_p) \wedge \forall x.(P(x) \supset \Phi(x))) \supset \forall x.(P(x) \equiv \Phi(x))). \qquad (14)$$

Or, semantically, *there exists no predicate Φ which is an element of A_p and which is a strict superset of P*:

$$\neg(\exists\Phi.((\Phi \in A_p) \wedge (P \subset \Phi))) \qquad (15)$$

The similarity, between the above default axiom and the predicate circumscription axiom allows us to take advantage of the techniques developed for circumscription in preceding sections. The default axiom, as it is stated above, however, is not yet adequate. Normally, a default is expressed about the property of a particular class of individuals in the domain, such as the default value of a slot for a particular frame [Bobrow 77], or the role filler of a concept [Brachman 85a]. Thus we must be able to restrict the application of a default to only those individuals that are subsumed by a particular concept. Using the results of section 8.1 the default axiom schema is restated as:

$$\forall\Phi.(((\Phi \in A_p) \wedge \forall x.((P(x) \wedge C(x)) \supset \Phi(x))) \supset \forall x.((\Phi(x) \wedge C(x)) \supset P(x))). \qquad (16)$$

Thus to state the default that "all birds fly" we specialize the above axiom schema with P as FLIES and C as BIRD:

$$\forall\Phi.((\Phi \in A_{\mathrm{FLIES}}) \wedge \forall x.((\mathrm{FLIES}(x) \wedge \mathrm{BIRD}(x)) \supset \Phi(x)) \qquad (17)$$
$$\supset \forall x.((\Phi(x) \wedge \mathrm{BIRD}(x)) \supset \mathrm{FLIES}(x))).$$

A nice property of the default axiom is that it allows us to specify *different defaults for different* classes of individuals, thus we might say that birds fly by default while humans don't. Furthermore, if we state that a set of properties are mutually exclusive and collectively exhaustive, then a default can

specify one of several slot fillers. For example we might want to say that all animals either swim, walk or fly, and then specify the defaults of each class of animals to be one of the three.

Next we need a way of determining if an individual, a, satisfies a default property, P. To do this we partially instantiate P with a, as described in section 7.2, except that we take $\Phi(x)$ to be $((x = a) \vee P(x))$. This produces the following axiom schema:

$$(((x = a) \vee P(x)) \in A_P) \wedge C(a)) \supset P(a) \tag{18}$$

Thus for the above example this schema becomes:

$$((((x = a) \vee \text{FLIES}(x)) \in A_{\text{FLIES}}) \wedge \text{BIRD}(a)) \supset \text{FLIES}(a)$$

Thus to determine whether an individual, a, satisfies a particular property P, we partially instantiate P with a in any schema for which 1) the class, C, subsumes a, and 2) the default is the property P.

For example, to determine that a snipe flies, given that "a penguin is a bird that cannot fly", we instantiate the above schema with a $\equiv$ snipe, and A $\equiv$ ($\text{BIRD}(\text{penguin}) \wedge \neg\text{FLIES}(\text{penguin})$):

$$(\text{BIRD}(\text{penguin}) \wedge \neg((\text{penguin} = \text{snipe}) \vee \text{FLIES}(\text{penguin}))) \wedge \text{BIRD}(\text{snipe}))$$
$$\supset \text{FLIES}(\text{snipe})$$

The first conjunct of the antecedent follows from what is known, thus, if a snipe is a bird then it can fly.

This completes the description of how to specify and use defaults. The last few sections have provided a number of extensions that greatly expand the scope of minimal reasoning to encompass several other forms of conjectural reasoning . In this section we have seen a number of these extensions come together in an analogous form of maximal reasoning, thus providing a precise semantics for default reasoning. Before continuing to another topic, it is worthwhile to consider for a moment the appropriateness of applying default and other conjectural techniques.

9.1. Should Defaults be the Default?

Although default reasoning and other rules of conjecture are essential components of the reasoning process, it is important that one does not become overzealous with their application. Default reasoning has been used liberally in a number of AI applications involving areas like knowledge representation, temporal reasoning and search.

In several applications there has been a tendency to perform blanket applications of defaults. For example, in the field of knowledge representation, Brachman [Brachman 85b] points out that if we remove the definitional import of a taxonomy and instead interpret all properties to be default properties, then it is no longer clear what the meaning of subsumption is. Our intuitions say that if

one concept is subsumed by another, then the subsumer places a set of necessary properties on the subsumee. Yet if we allow any property to be defaulted indiscriminantly, then this necessary condition no longer follows. In this case it is no longer clear what the semantics of a taxonomy is, and thus the integrity of the representation collapses.

In the area of temporal reasoning there has been a tendency to assume everywhere that the value of a quantity (position of a block, and so forth) doesn't change unless there is evidence to the contrary. The application of this assumption often seems to produce the right result and thus has appeared adequate. A serious problem, however, arises when one of these assumptions results in an inconsistency. If every fact in the system is an assumption, then any piece of knowledge relating to the inconsistency is suspect! Thus the system is presented with a vast number of alternatives to consider. Instead, by constructing more robust temporal representations it is often possible to deduce with complete confidence the duration of some event, while avoiding the cost of making (and possibly later retracting) any assumptions.

The point here is not that one should not use defaults or other rules of conjecture, since they play an essential role in many areas of reasoning. Instead it is a caution that one should be judicious about their use.

10. Other Forms of Circumscription

Predicate circumscription places a minimality condition on the set of individuals that have some property. It is useful to consider other places where a similar condition of minimality might be desirable. This section briefly discusses two such areas: 1) domains and 2) individuals. A third form of circumscription referred to as axiom circumscription is discussed in section 13.3.

In general the minimality condition used in circumscription consists of two parts. The first part is a test for membership in the set of things that are being minimized over, while the second provides a condition, based on an ordering relation for the set, that determines the elements of the set that are minimal. In predicate circumscription the test is logical consistency, the elements are extensions of predicates, and the ordering relation is subset. There are, however, several other ways in which one might want to use minimality. In each of the following cases the elements of the set we are trying to minimize, denoted A_{OK}, are sets and the ordering relation is proper subset. The minimality condition for an element S is then summarized as follows:

1. $S \in A_{OK}$ and

2. $\neg(\exists \Phi.((\Phi \in A_{OK}) \wedge (\Phi \subset S)))$

10.1. Domain Circumscription

Earlier we said that the notion that predicate circumscription captures is that a closed predicate is true of the smallest set of individuals which is consistent with what is known. Domain circumscription says that the set of all individuals in the world is the smallest set that is consistent with what is known. Thus predicate circumscription minimizes the set of individuals that have some property, while domain circumscription minimizes the set of all individuals in the domain. For example, consider the statement "Peter, Ramesh and Gerry are in the room." Applying predicate circumscription to the predicate "in the room", we infer that these are the only people in the room. On the other hand, from domain circumscription, we infer that these three people are the only individuals there are.

Domain circumscription is primarily useful in reasoning about universally quantified hypothetical statements (or their equivalents) based on experience. For example, I might use domain circumscription to say that, "it is my experience that all cats are friendly," and then use this fact to deduce that it is safe to pet a particular large black and yellow striped cat. (Note that this form of reasoning is non-monotonic.) After a close encounter with the tiger, we will no doubt believe that the statement "all cats are friendly" is false, and our beliefs about cats will have changed non-monotonically.

To define the precise semantics of *domain* circumscription we let A_D be the set of all possible domains that are consistent with what is known, and let ALL represent the minimal domain. The semantics of domain circumscription is then:

1. $\text{ALL} \in A_D$ and

2. $\neg(\exists\Phi.((\Phi \in A_D) \wedge (\Phi \subset \text{ALL})))$

To construct the corresponding axiom in predicate calculus we need predicates that describe the sets: ALL and A_D. To handle the first we take ALL to be a predicate that is true of all individuals in the domain ($\forall x.\text{ALL}(x)$). A domain, Φ, is consistent with what is known if it contains all the individuals referred to by A. According to McCarthy, the domain consist of 1) all the constants mentioned in A, 2) all the individuals resulting from the application of any function symbol, f, mentioned in A to one of the constants of 1), and 3) those constants denoted by induction schemas (e.g., P(0) and $P(n) \supset P(succ(n))$).

Finally, McCarthy [McCarthy 80a] describes a way of constructing a predicate formula, $\Phi \in A_D$, that is true just in case the domain Φ satisfies the above three conditions.[17] The resulting axiom is then:

---

[17] McCarthy defines the predicate $\Phi \in A_D$ to be $\text{Axiom}(\Phi) \wedge A^{\Phi}$. "Axiom($\Gamma$) is the conjunction of sentences $\Phi(a)$ for each constant a and sentences $\forall x\,(\Phi(x) \supset \Phi(f(x)))$." Thus, Axiom($\Phi$) covers conditions 1) and 2) above. A^{Φ} (called the relativization of A with respect to Φ) covers condition 3), and is formed by replacing each universal quantifier "$\forall x$." in A by "$\forall x\,(\Phi(x) \supset$" and each existential quantifier "$\exists x$." in A by "$\exists x\,(\Phi(x) \wedge$".

$$\forall\Phi.(((\Phi \subset A_{I}) \wedge \forall x.(\Phi(x) \supset \text{ALL}(x))) \supset \forall x.(\text{ALL}(x) \equiv \Phi(x))). \qquad (19)$$

10.2. Individual Circumscription

A second possible use of minimality we refer to as *individual circumscription*. Predicate circumscription concerns itself with the minimal set of individuals that satisfy a predicate P. Similarly, individual circumscription concerns itself with the minimal set of *predicates* that are true of an individual I. More precisely, individual circumscription states that *the set of all predicates P that are true of an individual I is the smallest set of predicates that are consistent with the known facts A.*

One example of the utility of individual circumscription appears in the domain of hardware troubleshooting. Suppose we want to determine that a certain component is working. By circumscribing all the properties of the components, we can say that the component works unless one of these properties is a member of a class of failure properties. In addition, one might use a technique similar (but not identical) to individual circumscription to circumscribe all instances of the two place predicate CONNECTED(x,y) that mention a particular node, N. This provides us with the set of all connections to N, and can be used, for example, when applying Kirchoff's Current Law to N.[18]

Let A_I be a set, each of whose elements is a set of predicates, such that each of the predicates being true of the individual I is consistent with what is known. In addition, let PROPERTY-OF-I represent the minimal set of predicates true of I, then the semantics of individual circumscription is:

1. $\text{PROPERTY-OF-I} \in A_I$ and
2. $\neg(\exists\Phi.((\Phi \in A_I) \wedge (\Phi \subset \text{PROPERTY-OF-I})))$

To construct the corresponding axiom in predicate calculus we need predicates that describe the sets: PROPERTY-OF-I and A_I. PROPERTY-OF-I is defined to be a predicate on predicates that satisfies the second order axiom, "For all predicates P, PROPERTY-OF-I is true of P, iff P is true of I"(i.e.,$\text{PROPERTY-OF-I}(P) \equiv P(I)$). Recall that A_I is a set, each of whose elements is a set of predicates P, such that P(I) follows from A (where A is a set of first order axioms). Taking Φ to be a set of predicates, then $\Phi \in A_I$ can be constructed as a predicate formula consisting of the conjunction of sentences, $P(I) \supset \Phi(P)$, for each predicate P in the set of first order axioms A. That is, $\Phi \in A_I$ if Φ contains every predicate P mentioned in A that is true of I. Individual circumscription is then stated as a second order axiom quantifying over all predicates P, and all predicates on predicates Φ, such that:

$$\forall\Phi.(((\Phi \in A_I) \wedge \forall P.(\Phi(P) \supset \text{PROPERTY-OF-I}(P))) \supset \forall P.(\text{PROPERTY-OF-I}(P) \equiv \Phi(P))). \qquad (20)$$

---

[18] Kirchoff's Current Law states that the sum of the currents into a node is zero, where there is a current associated with each terminal connected to the node. Individual circumscription tells us all things connected to the node.

In this section we have explored the application of the minimality condition developed for predicate circumscription to two other aspects of conjectural reasoning. A third technique, axiom circumscription, is explored later during the discussion about the ways that resource limited agents focus on relevant information. Each of these techniques allow a set of facts to be circumscribed along a different dimension: predicates, individuals, axioms and the domain itself. The next section examines the claim that circumscription is a form of non-monotonic reasoning.

11. Circumscription and Non-Monotonic Reasoning

McCarthy and others have described circumscription as a form of non-monotonic reasoning. In this section we examine the notion of monotonicity and how it relates to circumscription. McCarthy defines monotonicity as follows:

> If a sentence q follows from a collection A of sentences and $A \subset B$, then q follows from B. In the notation of proof theory: if $A \vdash q$ and $A \subset B$, then $B \vdash q$.

Thus, in a monotonic logic, if it follows from a set of sentences A that all cats are friendly, then this will still follow if an arbitrary set of sentences are added to A.

Predicate circumscription, however, defines a notion of inference different from traditional logics called circumscriptive inference ($\vdash_p$). We write $A \vdash_P q$ if the sentence q follows from the result of circumscribing P in A. This implies that q is true in all models of A that are minimal in P. Unlike the normal notion of inference, circumscriptive inference is non-monotonic. The reason for this is best seen through an example. In the original missionary and cannibal problem we used circumscription to determine that the boat was working, based on the fact that we had no evidence to the contrary. However, if after carefully examining the boat we saw that it had a gaping hole, then we would no longer believe that the boat was working. Thus, using circumscriptive inference, a statement that follows from a collection of facts A no longer follows when that collection is expanded to a larger set B -- in other words, circumscriptive inference is non-monotonic.

The argument for the non-monotonicity of circumscription can be somewhat deceptive. It is true that the notion of circumscriptive inference as it is defined above is non-monotonic; however, does it accurately reflect how circumscription is applied in first order logic (FOL)? The problem is that circumscriptive inference manages to sweep a small but very important axiom under the rug, namely the predicate circumscription axiom. If we let $C_{A,P}$ denote the axiom that results in circumscribing P in A, then we can rewrite circumscriptive inference in terms of regular inference as follows:

$$A \vdash_P q \equiv (A \wedge C_{A,P}) \vdash q$$

That is, q follows from A *and* the instantiation of the circumscription axiom schema with A and P. If

we then take B to be a superset of A and say that it is not the case that $B \vdash_P q$, then this is equivalent to saying that it is not true that $(B \wedge C_{B,P}) \vdash q$. Thus, although A is a subset of B, it is not the case that $(A \wedge C_{A,P})$ is a subset of $(B \wedge C_{B,P})$. That is, if we circumscribe A and then add more sentences to get B, then we must remove the old circumscription axiom $C_{A,P}$ before circumscribing the resulting set. Thus, from this viewpoint the argument that circumscription is non-monotonic is invalid.[19]

The reason this discussion is important is that it raises a more general issue, that of making assumptions *explicit*. The reason that circumscription appeared non-monotonic is that we were hiding away the axioms that distinguished circumscription in the first place. The problem with hiding these assumptions is that when an inconsistency arises in the state of our world knowledge we are not able to consider these assumptions as possible causes. In most cases it is exactly these assumptions that have proven to be faulty. In these cases we want to be able to examine our assumptions and make a conscious decision about what to do next.

12. Computing with Circumscription

One of the most significant contributions of circumscription is that it provides a precise formalization of an interesting form of conjectural reasoning. The importance of such a formalization is that it allows us to explore both the expressiveness and the computational properties of a particular aspect of common sense reasoning. The discussion thus far has been heavily weighted towards the analysis of circumscription's expressiveness. During this discussion we have seen how the notion of predicate circumscription and the more general notion of minimality encompasses a broad class of conjectural reasoning including negation by failure, default reasoning, and (according to Reiter) the closed world assumption. In addition I have identified two common kinds of questions answered using circumscription. The first consists of determining the set of all individuals that satisfy a circumscribed predicate, while the second involves determining the truth of a circumscribed predicate about a particular individual. Finally, we have seen how the notion of minimality used in predicate circumscription can also be used to formalize other, similar types of reasoning.

During this discussion, however, I have essentially ignored the computational cost involved in using circumscription. This section focuses on the computational aspect of circumscription by examining the ways in which circumscription is used and analyzing some of the computational costs involved. To give this topic proper justice would require a much more extensive presentation than I have time to give here; thus, I will concentrate on the major issues, pointing to other publications on more detailed

[19] No doubt the reason McCarthy calls circumscription a form of non-monotonic reasoning, rather than a non-monotonic logic, is precisely the fact that, although it might be viewed as being non-monotonic, it works completely within the framework of a monotonic first order logic.

points. The interested reader is directed to [Non-Monotonic 84] for a collection of some of the most recent work on this topic.

With few exceptions, intelligent agents in the real world are resource limited. Thus for circumscription to be of practical use it must be effectively computable. Before jumping into the details of circumscription's computational properties it is useful to reexamine for a moment its motivation.

The need for circumscription (and the other rules of conjecture discussed) arises from the need to reason based on incomplete information. Given an incomplete description of the set of individuals satisfying a particular property, circumscription is an intuitively satisfying assumption about how to complete this set based on a notion of "simplicity". What the circumscription axiom provides us with is a precise way of stating this assumption. Given a predicate P and a set of axioms that we want to circumscribe over, the circumscription axiom is typically used to determine the set of all individuals that satisfy the predicate. To accomplish this, one first constructs a predicate Φ describing a set of individuals and then uses the circumscription axiom to show that Φ is equivalent to P, (where P is the circumscribed predicate we're interested in). This, in turn, is accomplished by instantiating the circumscription axiom schema with a specific A, P and Φ, and then showing that the antecedent of the axiom follows from what is known. Thus the major steps in using circumscription are 1) selecting a predicate P to be circumscribed, 2) selecting a set of axioms, A, to circumscribe over, 3) generating Φ, and 4) showing that the antecedent of the instantiated axiom follows from what we know.

The selection of a predicate to be circumscribed over is based on the particular domain being worked in and the problem being solved. Often circumscription is invoked when trying to prove a statement involving a predicate that has several possible extensions. In this case the predicate is circumscribed with the goal of coming up with a unique extension.

In all of the examples provided by McCarthy the set of axioms that are circumscribed over is the set of all the axioms known. This is not surprising; McCarthy's paper focuses on dealing with incomplete information. At no point are resource limitations or other forms of incompleteness considered; rather it is assumed that the agents being modeled are logically complete. Thus it would only make sense that an agent would use all the known information available. Of course, a physically limited agent may not be able to consider all the information available to him, and thus must focus in on those axioms that appear "relevant" to the problem. This topic is discussed in detail in section 13.2.

Potentially the largest computational bottle neck in using circumscription is the selection of the predicate Φ. Circumscription is a very powerful mechanism, as McCarthy demonstrates in one of his examples where he shows that the induction axiom on the natural numbers is a special case of

circumscription. However, this expressive power can make circumscription very expensive to compute with in the general case. If, in searching for all minimal extensions of P, the circumscription axiom is used as the test part of a random generate and test process, then the generation of predicates Φ is equivalent to generating all possible predicate expressions! A number of researchers have recently studied the problem of selecting Φ while restricting the class of axioms to which circumscription is applicable. As discussed earlier, Reiter [Reiter 82] claims that, if the set of axioms A is restricted to be horn in P, then the closed world assumption is implied by circumscription. If in fact this is the case (a possible counter example to this claim appeared in section 8.3), then the technique of predicate completion developed by Clark [Clark 78] can be used to construct a Φ that is guaranteed to be equivalent to P. In addition heuristic techniques for dealing with cases where A is not horn in P have been recently discussed in [Naqvi 85]. The importance of this type of technique is that they produce an instantiation for Φ that is equivalent to P, while avoiding the costly generate and test process described above.

For many of the cases where circumscription is applied there exists a unique minimal extension for P. McAllester [McAllester ??] points out that the mu calculus provides a decision procedure for determining whether or not a set of axioms has a unique minimal extension for P [Park 76]. This provides a characterization of an important class of axioms and warrants further investigation.

Once the circumscription axiom is instantiated, the remaining step is to show that the axiom's antecedent (i.e., $(\Phi \in A_p) \wedge \forall x.(\Phi(x) \supset P(x))$) logically follows from what is known. For resource limited agents it is not possible, except in restricted domains, to deduce all logical consequences from the facts at hand. Thus even in those cases where the antecedent follows, the agent may give up before proving that it is true. This issue is raised in section 14 during the discussion of Konolige's work on resource limitations.

If, instead of trying to completely characterize P, circumscription is used to determine the truth of P for a particular individual, then the technique proposed in section 7.2 is particularly useful, since it avoids the need to search for an instantiation of Φ. By using this technique the proof that P(a) is false is equivalent to proving the axiom resulting from substituting every occurrence of P(x) in A with $((x \neq a) \wedge P(x))$. This is much simpler than first going through the random generate and test process described above to find an instantiation for Φ.

This section has examined a number of the major computational costs incurred when using circumscription. At one point during the discussion I mentioned the fact that the selection of the set of axioms, A, that are being circumscribed depends on the necessity of a resource limited agent to focus his attention. Focusing one's attention involves making deductions based only on a small set of

relevant facts, while ignoring any facts irrelevant to the problem. In order to formalize the notion of "focusing one's attention" the logic formalism must provide a means of describing the truth derivation process, that is, the derivability of a fact from a specific set of axioms. Konolige provides such a formalism. The properties of this formalism and its applicability to various forms of conjectural reasoning is the topic of the next section.

13. Relevance Incompleteness

McCarthy's work on circumscription focuses on the problem of making decisions based on incomplete knowledge of the world. In addition to this problem, intelligent agents in the "real" world must cope with several other forms of incompleteness due to physical limitations. The focus of Konolige's work [Konolige 84], [Konolige 82] is on a formal system that can be used to model several of these limitations. Konolige addresses three forms of incompleteness which he refers to as: 1) relevance incompleteness, 2) resource limited incompleteness, and 3) fundamental logical incompleteness. Relevance incompleteness occurs when an agent has available all the necessary information to deduce the desired consequences, but restricts his set of knowledge in such a way that the deduction is no longer possible. Resource limited incompleteness occurs when ". . . an agent has the inferential capabilities to derive some consequence of his beliefs but simply does not have the computational resources to do so." Finally, fundamental logical incompleteness occurs when an agent has a logically incomplete or inconsistent inference procedure. Relevance incompleteness is explored below, while the other two forms of incompleteness are explored briefly in the next section.

The life of a typical individual is cluttered with an incredible number of inconsequential facts, far too many to cope with as a single body of knowledge when solving any sizable problem. The treasured piece of knowledge about the song on the second track of the flip side of a Supertramp album, which was so helpful during last night's trivial pursuit game, is not going to do one bit of good during the next morning's calculus exam. Thus one must be able to determine what information is relevant for a particular situation. For example in the missionary and cannibal problem, one immediately determines that knowledge about boats, rivers and transportation is relevant, while last night's dinner and the mating patterns of overweight penguins are not. Thus to solve the missionary and cannibal problem we "circumscribe" a collection of facts that appear useful, considering all others to be irrelevant to the problem at hand. Konolige provides a formalization of this idea which he refers to as *circumscriptive ignorance*.

13.1. Circumscriptive Ignorance

Predicate circumscription is used to draw a circle around a set of facts that are believed to be true, considering anything outside of the circle to be false. Similarly using circumscriptive ignorance, a circle is drawn around a set of facts such that any facts outside of the circle are considered to be *irrelevant* to the problem. Thus, if the desired result cannot be deduced from the set of circumscribed facts, then it is assumed that the result is not deducible from any larger set. More precisely, let B be a set of axioms that are known ("believed") and let A be a subset of B that are considered to be relevant. By circumscribing the relevant facts A when deducing P we say that, if P does not follow from A then P does not follow from B. For example, circumscriptive ignorance might be used to say something like, "At night, if I can't find a missing needle under the street light, then I can't find it at all," or "If he can't figure out the riddle with all the clues he's been given then he is just not going to be able to figure it out."

To formalize circumscriptive ignorance we must be able to make an explicit statement about the derivability of a fact P. Furthermore, it is necessary to be able to specify a particular set of facts that P is being derived from. To accomplish this Konolige introduces what he refers to as the *circumscription operator*. The operator I will use in this discussion (denoted by angle brackets) is a simplified version of Konolige's operator, which ignores for the moment the issues of 1) resource limitations, and 2) modeling the beliefs of multiple agents. The issue of resource limitations is addressed in the next section. The intended meaning of the circumscriptive atom, $\langle\Gamma\rangle P$, is that the sentence P follows from the set of sentences Γ, that is $\Gamma \vdash P$. Similarly, $\neg(\langle\Gamma\rangle P)$ means that P does not follow from Γ. It is important to note that the circumscriptive atom, $\langle\Gamma\rangle P$, is only semi-decidable using the axiomatization of first order predicate calculus. Thus proving $\neg(\langle\Gamma\rangle P)$ requires that P is not derivable from Γ, an undecidable question. In reference [Konolige 82], the circumscription operator is incorporated into a propositional modal logic based on Sato's K4 [Sato 76], a logic that has been proven to be decidable. Taking B to be a base set of axioms representing what we know, and A to be the set of axioms that are being circumscribed over, then circumscriptive ignorance is expressed as:

$$\langle B\rangle A \supset (\langle A\rangle P \equiv \langle B\rangle P)$$

In the forward direction this says that, given that the relevant set of facts A is a subset of what we know (A follows from B), then if the desired fact P follows from A, then P also follows from what we know. This is certainly true for any monotonic logic, since anything that follows from a set of axioms also follows from a superset of those axioms. The reverse direction is more interesting and says that (again given that A follows from B) if P cannot be inferred from A, then P cannot be inferred from what we know. This statement differs from the formalization provided by Konolige in that it requires A to be a subset of what is known "$\langle B\rangle A \supset \ldots$". Without this restriction it is possible to select a set of facts

A that are disjoint from what we know. The interpretation of circumscriptive ignorance then would be, "if P cannot be inferred from a set of facts, which we possibly don't know about, then P cannot be inferred from what we know". This, however, is not the desired semantics.

The notion of circumscriptive ignorance is very similar to that of failure by negation [Clark 78]. Failure by negation says that if P cannot be inferred from what is known then P is false. Using Konolige's circumscription operator in a manner analogous to circumscriptive ignorance, we can restrict the proof of the truth of P to a subset of what we know. Failure by negation then says that if P cannot be inferred from A (a subset of what we know) then P is false. Thus, the difference between the two techniques is that failure by negation makes the stronger conclusion that P is false, while circumscriptive ignorance says that P cannot be inferred. Negation by failure can also be used (as it is in the closed world assumption) to say that the only facts that are true are those that follow from the facts A. By analogy, circumscriptive ignorance can be used to say that those facts that are derivable from A are all that can be derived.

The ability, provided by the circumscription operator, to make explicit statements about the truth derivation process played an essential role in formalizing circumscriptive ignorance. The circumscription operator also allows us to formalize several other forms of conjectural reasoning not possible in standard first order logic. Two such rules of conjecture are the topics of the next two sections.

13.2. The Relevant Axioms: Restricting the Scope of A in Circumscription

In section 8 we examined ways of focusing the circumscription axiom on the relevant set of predicates and individuals in the domain. In this section we examine a way of focusing on the relevant set of axioms to which circumscription is applied.

As was mentioned in section 12, when predicate circumscription is being applied by an "omnipotent" agent, the selection of A, the facts that are being circumscribed is the set of all things that the agent knows about. In other words, the agent will want to use all of the available information to reason with. On the other hand, a resource limited agent, often cannot afford the cost of wading through all the information available. In addition there may be some additional cost incurred in acquiring the knowledge, such as searching through reference libraries or performing experiments. Thus, in these situations the agent must restrict the set of axioms, A, to those that he considers relevant, based on properties of the particular problem.

Examining the predicate circumscription axiom of equation (1) we note that, although it is possible to select any set of axioms to be circumscribed over, it is not possible to restrict the set of axioms

used to prove the antecedent of the circumscription axiom. Thus, we are restricting the set of axioms A used to test the consistency of the circumscribed predicate P, thereby reducing our confidence that P is correct. Yet at the same time we are still burdened with the cost of considering all of our knowledge in proving that our "guess" for P is minimal (that is, $\Phi \equiv P$). In other words we get the worst of both worlds: a weakening in logical completeness, without any gain in computational efficiency.

Recall that the reason for restricting the set of facts A to a subset of what is known is to reduce the computational burden of using circumscription in the first place. Thus we would like to assume that the facts A are the only facts relevant to the problem and completely ignore everything else we know in applying predicate circumscription. Konolige provides us with exactly the right framework to accomplish this. To achieve the desired result Konolige's circumscription operator is used to restrict the deduction of the antecedent of the circumscription axiom so that it is true only if it follows from the relevant axioms A. This produces an axiom that combines McCarthy's predicate circumscription and Konolige's circumscriptive ignorance:

$$\forall \Phi.(\langle A \rangle((\Phi \in A_p) \wedge \forall x.(\Phi(x) \supset P(x))) \supset \forall x.(P(x) \supset \Phi(x))). \tag{21}$$

Thus the above axiom changes the meaning of predicate circumscription from "it follows from what we know that P is minimal in A," to "it follows from A that P is minimal in A"

13.3. Axiom Circumscription - Communicating Ideas Effectively

In earlier sections I have discussed three types of circumscription: 1) Predicate circumscription, which minimizes the set of individuals that have some property, 2) Domain circumscription, which minimizes the set of individuals in the domain and 3) Individual circumscription, which minimizes the set of properties that a particular individual has. Using Konolige's circumscription operator I now introduce a fourth type of circumscription, related to the notion of a minimal set of axioms used to deduce a particular fact. Due to an agent's physical limitations, there is a computational cost incurred when an intelligent agent makes a set of deductions. If this agent interacts with other intelligent agents, then there is an additional cost due to the finite bandwidth of the communication channel. The tradeoff between these two costs is discussed in section 14. In this section I assume that the cost of making deductions is negligible, and instead focus on the problem of minimizing the communication cost.

One problem in communicating an idea is that there is a huge amount of negative information that, if stated explicitly, will be very expensive to communicate. This issue is of great concern to researchers studying database systems [Gallaire 78], and was the topic of section 7.3.

A second problem involves the cost in communicating a sufficient amount of information such that the desired meaning is conveyed. To minimize communication costs, while conveying a particular set of ideas, we would like to transmit the minimum amount of *information necessary to infer the ideas.* (i.e., we would like to eliminate irrelevant information). Notice in the previous sentence that the magic words, *minimize* and *irrelevant* appear once again. To formalize the idea of minimizing transmitted information, I draw both from the ideas of McCarthy on circumscription and the ideas of Konolige on relevance. I will refer to the formalization of this type of conjecture as *axiom circumscription.*

Let P be a set of ideas that we would like to convey and let A be the set of facts that we communicate in trying to convey P Then our goal is to provide a formal statement of the idea: *"The set of facts A, used to communicate the idea P is minimal."* To formalize this statement I need to define 1) what it means for "facts A to communicate the idea P" and 2) what it means to be minimal. I will take the statement "the facts A communicate the idea P" to mean that the sentence P can be inferred from A, or using Konolige's notation $\langle A \rangle P$. Next I use the notion of minimality discussed in section 10, i.e., S is minimal in A_{OK} if:

1. $S \in A_{OK}$ and

2. $\neg(\exists\Phi.((\Phi \in A_{OK}) \wedge (\Phi$ "is strictly smaller than" $S)))$

A_{OK} is taken to be the set of all axioms that 1) follow from what is known, and 2) can be used to derive S. Thus the condition for the set of axioms Φ being an element of A_{OK} is stated as "$\Phi \wedge \langle\Phi\rangle S$." In addition, a set of axioms B is smaller than a second set Γ, if B follows from Γ (i.e., $\langle\Gamma\rangle B$), and B is strictly smaller than Γ if, in addition, it is not the case that Γ follows from B (i.e., $\langle\Gamma\rangle B \wedge \neg\langle B\rangle\Gamma$). Finally, we can construct an axiom, similar to that of predicate circumscription, that captures this semantics (where Φ is quantifying over all sentences):

$$\forall\Phi.((\Phi \wedge \langle\Phi\rangle S \wedge \langle A\rangle\Phi) \supset (A \equiv \Phi)). \tag{22}$$

The axiom above states that any set of known axioms Φ that can be used to deduce S and that follows from A is equivalent to A. Adding the restriction that A is a known set of axioms such that S can be deduced from A (i.e., $A \wedge \langle A\rangle S$), then *A is the minimal set of known axioms that S can be deduced from.*

One thing this axiom doesn't take into account is the fact that the agent being communicated with already has some knowledge of his own. We would like to avoid restating all of his world knowledge when communicating our idea. Taking B to be the knowledge of this second agent, then axiom circumscription becomes:[20]

---

[20] A better statement of the axiom circumscription axiom would include the use of Konolige's belief operator to explicitly describe the knowledge of the two communicating agents.

$$A \wedge \langle A \wedge B \rangle S \tag{23}$$

$$\forall \Phi . ((\Phi \wedge \langle \Phi \wedge B \rangle S \wedge \langle A \wedge B \rangle \Phi) \supset (A \equiv \Phi)).$$

where the first equation says that S follows from A taken together with B, and the second equation says that A is the minimal set of axioms satisfying this condition. Axiom circumscription conveys the idea that a set of sentences A is logically compact, that is, there is no set of sentences (not logically equivalent to A) that convey only a subset of the ideas that A conveys and yet convey the desired set of ideas S. This idea occurs over and over again in real life. For example, when writing a paper, the author will iterate through several drafts in the pursuit of a document that states clearly and concisely the desired meaning. If this author fails to achieve this goal, and the ideas are sufficiently significant, then other authors (or students taking area exams), will pick up the task.

Of course, there is still much that this axiom doesn't capture. For example, the natural numbers can be defined using an induction schema or by explicit enumeration of an infinite number of individuals. It is clear that the induction schema is a more concise description, yet the two are logically equivalent, and thus would be considered equivalent from the viewpoint of axiom circumscription. Thus conciseness must take into account the cost in physically transmitting an idea, which is more closely linked to the syntax of the sentences used to capture the idea than the meaning of these sentences.

In addition, in the above scenario, the author's goal included *clarity*, as well as conciseness. Many advanced graduate mathematics texts have been written that are concise and convey the desired meaning, and yet are incredibly difficult for most people to read. For example, we might say that a concise statement of the majority of electronics is the set of Maxwell's equations. The problem here is that the above axioms do not account for the cost of the agent on the receiving end being able to derive the desired meaning from the set of sentences that have been communicated. Fixing this problem does not simply involve changing a few terms in the above axiomatization. The reason is that first order logic provides no means of modeling the cost of an inference, thus it is impossible to formalize the notion of minimizing the computation involved in "understanding" the idea being communicated. The problem of modeling forms of resource limited reasoning similar to this is the topic of the next section.

14. Resource Limited Incompleteness

Suppose for a moment that you are a freshman and it is your first day of classes. As you sit in a large lecture hall for your first class, your math professor walks in, writes a few fundamental axioms on the black board, and then a moment later announces that he has finished teaching you the course material for that semester and class is dismissed. Much to your surprise (and delight) the professor

also announces that the final exam has been canceled since he is sure that everyone in attendance now has a firm grasp of the material and its consequences.

This scenario seems preposterous to most of us, yet the actions of the professor are consistent with a model of reasoning based on first order logic. The problem with this model is that it does not take into consideration any of the many types of limitations of physical agents. In this section I summarize Konolige's efforts at avoiding this mistake in his formalism, focusing primarily on its relevance to circumscriptive reasoning. (A full account of this topic is beyond the scope of this paper as fit would appear that this author is somewhat resource limited).

Our friend the professor suffers a number of problems similar to those of more traditional logic systems. The reason that he wrote only a few lines on the blackboard is that he assumed one could easily deduce all logical consequences of what he wrote, that is he assumed that the students were logically complete. Furthermore, he assumed that everyone in the classroom could perform these deductions instantaneously. Finally, there was no need to test the students since it was assumed that their inference rules were logically sound. It is clear to most of us that humans do not think this way. Thus when trying to provide a formal theory modeling physically limited agents it is important that the theory takes into account these limitations. Konolige identifies two forms of incompleteness related to this discussion. The first is referred to as resource limited incompleteness and occurs when ". . . an agent has the inferential capabilities to derive some consequence of his beliefs but simply does not have the computational resources to do so." The second limitation he refers to as fundamental logical incompleteness, which occurs when an agent has a logically incomplete or inconsistent inference procedure.

AI has taken basically two approaches to the problem of dealing with resource limitations. The first approach involves weakening the expressive power of the language used to model a portion of the agent's reasoning process. An example of this is Krypton [Brachman 83], a knowledge representation system developed by Brachman and Levesque. One of the results of this research effort was to characterize a number of restricted definitional languages. During this research it became quite apparent that performing inference with even seemingly simple languages was computationally intractable (e.g., np-complete). [Brachman 84] However, the languages that have proven to be effectively computable (and even those that haven't) have not been sufficiently expressive for most domains.

A second approach taken by AI has been to weaken the inference power of a system while leaving the expressive power of the language intact. A simple example of this are chess programs with some number of levels of look ahead. A second example is the THNOT operator in micro-planner [Sussman

70] which roughly said: *Try to prove A. If you fail, assume not A is proved.* This is very similar to failure by negation, except that the test for failure has been weakened from, "does not logically follow," to "cannot be proven by micro-planner." The problem with THNOT is that there exists no statement of what it means for micro-planner to prove something, other than reading the code itself.

The approach taken by Konolige is not really an approach at all, but a framework in which to explore either of the two approaches discussed above. Konolige provides a formal syntactic system with the goal of being able to describe precisely a wide class of systems like micro-planner, and thus be able to make statements about their logical and computational characteristics. The contribution of Konolige's formalism is that it allows one to make explicit statements about logical incompleteness and belief.

14.1. Consequential vs Derivational Closure

To account for logical incompleteness, Konolige replaces consequential closure with the weaker constraint of *derivational closure*. Konolige points out that logical consequence is a semantic notion which states that A is a logical consequence of B if B holds in all models of A. On the other hand, derivational consequence is a syntactic notion about the ability to derive a fact from another fact and a set of syntactic rules. In Konolige's formalism an agent is modeled by a *deduction structure* consisting of a set of sentences representing the agents base beliefs, and a set of deduction rules that operate on these base beliefs to construct other beliefs. For an agent to be derivationally closed means that any sentence derivable from the initial set of beliefs using the deduction rules is also a member of the agent's beliefs.

The motivation for relaxing the constraint of consequential closure is clear: physically limited agents cannot deduce all logical consequences of everything they know. The motivation for derivational closure is less clear. Konolige states that "the chief motivation for requiring derivational closure is that it simplifies the technical task of formalizing the deduction model." It is rather unfortunate that Konolige provides no compelling examples of how this new form of closure simplifies the formalization process, and thus must be taken on faith.

One thing that derivational closure does provide is the ability to model a wide class of systems. In those cases where the set of rules are shown to be logically complete, the notions of consequential closure and derivational closure are equivalent, and all the problems of decidability are inherited. At the other end of the spectrum, if a system is provided with no deduction rules then the system becomes regular syntactic, that is, an agent believes a fact only if it is a member of its base beliefs. This provides an appropriate model for a simple database query system based on syntactic retrieval.

14.2. Derivational Closure and the Formalization of Conjecture

As discussed in the previous section, Konolige provides an operator, the circumscription operator, that makes it possible to talk about the derivation process explicitly. The effect that derivational closure has on this operator is that the operator no longer refers to logical consequence, but instead makes statements about derivability from a base set of rules. Thus something may be a logical consequence of something else and yet not be derivable. Furthermore, if the inference rules are not sound, then it may be possible to derive something that is not a logical consequence of what is known.

It is important for several of the rules of conjecture discussed earlier to be able to make explicit statements about the derivation process. For example, we can describe THNOT using the same statement as the one used to describe negation by failure. That is, letting B represent what we know, both rules are equivalently stated as:

$$\neg\langle B\rangle P \supset \neg P.$$

These two techniques are then distinguished by the set of rules used to model the agent. For example, using a set of axioms that are logically complete provides us with a formalization of failure by negation, while constructing a set of axioms that describe micro-planner provides a formalization of THNOT.

The interpretation of the circumscription operator to mean derivability as opposed to logical consequence also places a more realistic interpretation on axiom circumscription. Recall that axiom circumscription said roughly that the set of sentences A we want to use to communicate an idea P to an agent with knowledge B, is the minimal set of sentences that P logically follows from. This statement, however, did not take into consideration the derivability of P. Thus based on this statement, the scenario where the professor conveyed the knowledge of a math course by stating a few fundamental axioms would be perfectly reasonable.

On the other hand, using the circumscription operator to refer to derivation, as opposed to logical consequence, axiom circumscription acquires the more desirable interpretation that A is the minimal set of sentences used to *derive* P by the agent being communicated to. If the agent being talked to is a child then the information communicated to him will be sufficiently close to P that the child is required to make very few deductions in determining P. On the other hand, if the conversation is technical and the agent being communicated to is well versed in the field, then it may be adequate for A to be a single phrase.

The primary effect that the weakening of consequential closure has on predicate circumscription is that it no longer guarantees that we will be able to show that a predicate Φ is equivalent to the minimal

extensions of P in A. The reason for this is that, even though the antecedent of the circumscription axiom may follow from what is known (and thus the equivalence stated in the axiom's consequence), it is not necessarily the case that the antecedent will be derived. In this case the promise of Konolige's formalism is that it will provide a framework in which to characterize those cases when circumscription can be confidently applied, given a particular model of resource limitations. Under consequential closure the minimality condition used in predicate circumscription is based on the set of predicates consistent with what follows from A. Under derivational closure the minimality condition can instead be based on the set of predicates consistent with what is *derivable* from A. Thus Konolige's formalism provides us with a final dimension along which we can focus the circumscription axiom, that is, the proof derivation process.

This also suggests an iterative, heuristic technique for computing the circumscription of P, given that the logic is derivationally closed. The technique essentially starts with an initial guess for the circumscribed P and then goes through a relaxation process to find a minimal predicate. A reasonable guess for P might be constructed using a technique like Clark's predicate completion discussed earlier. In some cases, where the data base is not horn in P, Clark's technique will produce an extension of P that is inconsistent with what can be derived. If this occurs then, using information gained from the inconsistency, the extension of P is expanded to include one more individual and the deduction rules are used to test for consistency. This process is continued until the set becomes consistent with what is known, at which point a minimal extension has been founded. This process is repeated with each possible augmentation of the initial guess to find all minimal extensions of P.[21] On the other hand, if the initial guess starts out being consistent, then the process moves in the other direction by reducing extensions until an inconsistency is found. Given the fact that the deduction rules may not be logically complete, the agent may not be able to recognize an inconsistency when it occurs. Thus using this heuristic may produce an extension for P that is smaller than the actual minimal extension.

15. Conclusion

In this paper I have discussed a number of ideas related to the problem of modeling the incompleteness of intelligent agents in the physical world. During this discussion I have focused on the work of McCarthy on formalizing rules of conjecture for dealing with incomplete information, and the work of Konolige on modeling resource limited reasoning. McCarthy uses the notion of minimality to capture the intuition that "the set of things that have a particular property is the smallest set that is

---

[21] It is also necessary to have some means, of determining how many minimal extensions P has. As discussed earlier, the mu calculus [Park 76] may provide such a technique for determining when P has a single extension.

consistent with what is known." This intuition is captured by a formal rule of conjecture, predicate circumscription, which is expressed as an axiom of first order logic. This formalism is particularly desirable in that it works within the framework of an existing logical system that is well understood. On the other hand, to model resource limited reasoning Konolige must modify the formalism of traditional logics. This modification takes on two forms: 1) the relaxation of the constraint that the logic be consequentially closed, and 2) the addition of a set of modal operators for making explicit statements about belief and the proof derivation process. The expressive power of this new logical system is demonstrated by formalizing the notion of circumscriptive ignorance -- if something is not derivable from a set of relevant facts then it is not derivable.

In this paper I have analyzed the semantics of McCarthy's predicate circumscription, Konolige's circumscriptive ignorance, and several related formalisms, such as negation by failure, the closed world assumption, default reasoning and THNOT. In addition I have extended circumscription along several dimensions. First, the predicate circumscription axiom was modified to allow the ability to focus on a particular set of relevant predicates, individuals, and axioms to be circumscribed over. This is accomplished by, a) restricting the set of individuals that are being circumscribed over, b) expanding the number of predicates circumscribed and c) restricting the set of axioms used in performing the circumscription. Second, the notion of maximality (the inverse of minimality used in predicate circumscription) was used to formalize default reasoning. Third, the concepts of minimality and relevance were used to describe several novel forms of conjectural reasoning. Finally, predicate circumscription, as well as several other rules of conjecture, were extended, using Konolige's circumscription operator, to account for resource limitations. This provides a synthesis between the formalisms of McCarthy and Konolige applied to conjectural reasoning.

16. Acknowledgments

This is a subset of a paper originally written for my area exam. I would like to thank the members of my area exam committee, Gerry Sussman, Ramesh Patil and Peter Szolovits, who provided the primary impetus behind this paper. In addition I would especially like to thank Dave McAllester, Johan de Kleer, Hector Levesque, Danny Bobrow, Mark Shirley, Walter Hamscher, Jeff Van Baelen, Dan Weld and Leah Williams for many helpful comments and discussions.

References

[Bobrow 77]
Bobrow, D. G., and Winograd, T.
An Overview of KRL, a Knowledge Representation Language.
Cognitive Science 1(1):3-46, 1977.

[Boyer 79]
Boyer, R. and Moore, J. S.
A Computational Logic.
Academic Press, New York, 1979.

[Brachman 83]
Brachman, R. J., Fikes, R. E., Levesque, H. J.
Krypton: A Functional Approach to Knowledge Representation.
IEEE Computer, Special Issue on Knowledge Representation, 1983.

[Brachman 84]
Brachman, R. J.and Levesque, H. J.
The Tractability of Subsumption in Frame-Based Description Languages.
In *National Conference on Artificial Intelligence*. AAAI, August, 1984.

[Brachman 85a]
Brachman, R. J., and Schmolze, J. G.
An Overview of the KL-ONE Knowledge Representation System.
Cognitive Science 9(2):171-216, 1985.

[Brachman 85b]
Brachman, R. J.
'I Lied About The Trees' (or, Defaults and Definitions in Knowledge Representation).
The AI Magazine 6(3), 1985.

[Clark 78]
Clark, K. L.
Negation as Failure.
In Herve Gallaire and Jack Minker, editor, *Logic and Databases*. Plenum Press, New York, 1978.

[Collins 75]
Collins, A.M., et. al.
Reasoning from Incomplete Knowledge.
In Bobrow, D. G.and Collins, A. M., editor, *Representation and Understanding*, pages 383-416. Academic Press, New York, 1975.

[Doyle 80]
Doyle, J.
A Model for Deliberation, Action, and Introspection.
Technical Report TR 581, MIT Artificial Intelligence Laboratory, 1980.

[Gallaire 78]
Herve Gallaire and Jack Minker.
Logic and Databases.
Plenum Press, New York, 1978.

[GODS ??]
The Gods Must Be Crazy.
at the Orson Welles Theater

[Haack 78]
Haack, S.
Philosophy of Logics.
Cambridge University Press, Cambridge, 1978.

[Hayes 77]
Hayes, P.J.
In Defence of Logic.
In *Fifth International Joint Conference on Artificial Intelligence.* IJCAI, 1977.

[Konolige 82]
Konolige, K.
Circumscriptive Ignorance.
In *National Conference on Artificial Intelligence.* AAAI, August, 1982.

[Konolige 84]
Konolige, K.
Belief and Incompleteness.
Technical Report TR 319, SRI International, January, 1984.

[Levesque 84]
Levesque, H. J.
A Logic of Implicit and Explicit Belief.
In *National Conference on Artificial Intelligence.* AAAI, August, 1984.

[McAllester ??]
McAllester, D.
personal communication

[McCarthy 77]
McCarthy, J.
Epistomelogical Problems of Artificial Intelligence.
In *Fifth International Joint Conference on Artificial Intelligence.* IJCAI, 1977.

[McCarthy 80a]
McCarthy, J.
Circumscription--A Form of Non-Monotonic Reasoning.
Artifical Intelligence 13:27-39, 1980.

[McCarthy 80b]
McCarthy, J.
Addendum: Circumscription and other Non-Monotonic Formalisms.
Artifical Intelligence 13:171-172, 1980.

[McDermott 80]
McDermott, D and Doyle, J.
Non-Monotonic Logic I.
Artifical Intelligence 13:41-72, 1980.

[Mendelson 64]
Mendelson, E.
Introduction to Mathematical Logic.
D. Van Nostrand Company, Inc., Princeton, New Jersey, 1964.

[Naqvi 85]
Naqvi, S. A.
A Characterization of Negative Knowledge in Reasoning Systems.
In Michael Brodie, editor, *Proceedings of the Islamorada Workshop on Large Scale Knowledge Base and Reasoning Systems.* February, 1985.

[Non-Monotonic 84]
Non-Monotonic Reasoning Workshop
New Paltz, NY 12561, 1984.

[Park 76]
Park, D.
Finiteness is mu-ineffable.
Theoret. Comput. Sci. 3:173-181, 1976.

[Reiter 78]
Reiter, R.
On Closed World Data Bases.
In Herve Gallaire and Jack Minker, editor, *Logic and Databases.* Plenum Press, New York, 1978.

[Reiter 80]
Reiter, R.
A Logic for Default Reasoning.
Artifical Intelligence 13:81-132, 1980.

[Reiter 82]
Reiter, R.
Circumscription Implies Predicate Completion (Sometimes).
In *National Conference on Artificial Intelligence.* AAAI, August, 1982.

[Sato 76]
Sato, M.
A Study of Kripke-type Models for Some Modal Logics by Gentzen's Sequential Method.
Technical Report, Research Institute for Mathematical Sciences, Kyoto University, July, 1976.

[Suppes 57]
Suppes, P.
Introduction to Logic.
Van Nostrand Reinhold Company, New York, 1957.

[Sussman 70]
Sussman, G., Winograd, T. and Charniak, E.
Micro-planner reference manual.
MIT, Cambridge,MA, 1970.
AIM-203(1970)

[Weyhrauch 80]
Weyhrauch, R.
Prolegomena to a Theory of Mechanized Formal Reasoning.
Artifical Intelligence 13:133-170, 1980.

[Winograd 80]
Winograd, T.
Extended Inference Modes in Reasoning by Computer Systems.
Artifical Intelligence 13:5-26, 1980.

END

DTIC

5-86